Uplift Modeling with ROC: An SRL Case Study

Houssam Nassif, Finn Kuusisto, Elizabeth S. Burnside, and Jude Shavlik

University of Wisconsin, Madison, USA

Abstract

Uplift modeling is a classification method that determines the incremental impact of an action on a given population. Uplift modeling aims at maximizing the area under the uplift curve, which is the difference between the subject and control sets' area under the lift curve. Lift and uplift curves are seldom used outside of the marketing domain, whereas the related ROC curve is frequently used in multiple areas. Achieving a good uplift using an ROC-based model instead of lift may be more intuitive in several areas, and may help uplift modeling reach a wider audience.

We alter SAYL, an uplift-modeling statistical relational learner, to use ROC instead of lift. We test our approach on a screening mammography dataset. SAYL-ROC outperforms SAYL on our data, though not significantly, suggesting that ROC can be used for uplift modeling. On the other hand, SAYL-ROC returns larger models, reducing interpretability.

SAYL-ROC

SAYL is a Statistical Relational Learner based on SAYU that integrates uplift modeling with the search for relational rules. Similar to SAYU, every valid rule generated is used to construct a Bayesian network (alongside with current theory rules) via propositionalization, but instead of constructing a single classifier, SAYL constructs two TAN classifiers; one Bayes net for each of the subject and control groups. Both classifiers use the same set of attributes, but are trained only on examples from their respective groups. SAYL uses the TAN generated probabilities to construct the lift and uplift curves, where area under the uplift curve (AUU) is the difference in areas under the lift curves (AUL).

$$AUU = AUL_s - AUL_c = \Delta(AUL)$$

If a rule improves $\Delta(AUL)$ by threshold θ, the rule is added to the attribute set. Otherwise, SAYL continues the search.

We implement SAYL-ROC, a SAYL variant that computes area under the ROC curve (AUC) instead for each of the groups using the two classifiers, and returns $\Delta(AUC)$ as a rule score to guide the search. SAYL thus optimizes for $\Delta(AUL)$, while SAYL-ROC optimizes for $\Delta(AUC)$.

Algorithm 1 SAYL

$Rs \leftarrow \{\}$; $M_0^s, M_0^c \leftarrow InitClassifiers(Rs)$
while $DoSearch()$ **do**
  $e_s^+ \leftarrow RandomSeed()$;
  $\perp_{e_s^+} \leftarrow saturate(e)$;
  while $c \leftarrow reduce(\perp_{e_s^+})$ **do**
    $M^s, M^c \leftarrow LearnClassifiers(Rs \cup \{c\})$;
    if $Better(M^s, M^c, M_0^s, M_0^c)$ **then**
      $Rs \leftarrow Rs \cup \{c\}$; $M_0^s, M_0^c \leftarrow M^s, M^c$;
      break
    end if
  end while
end while

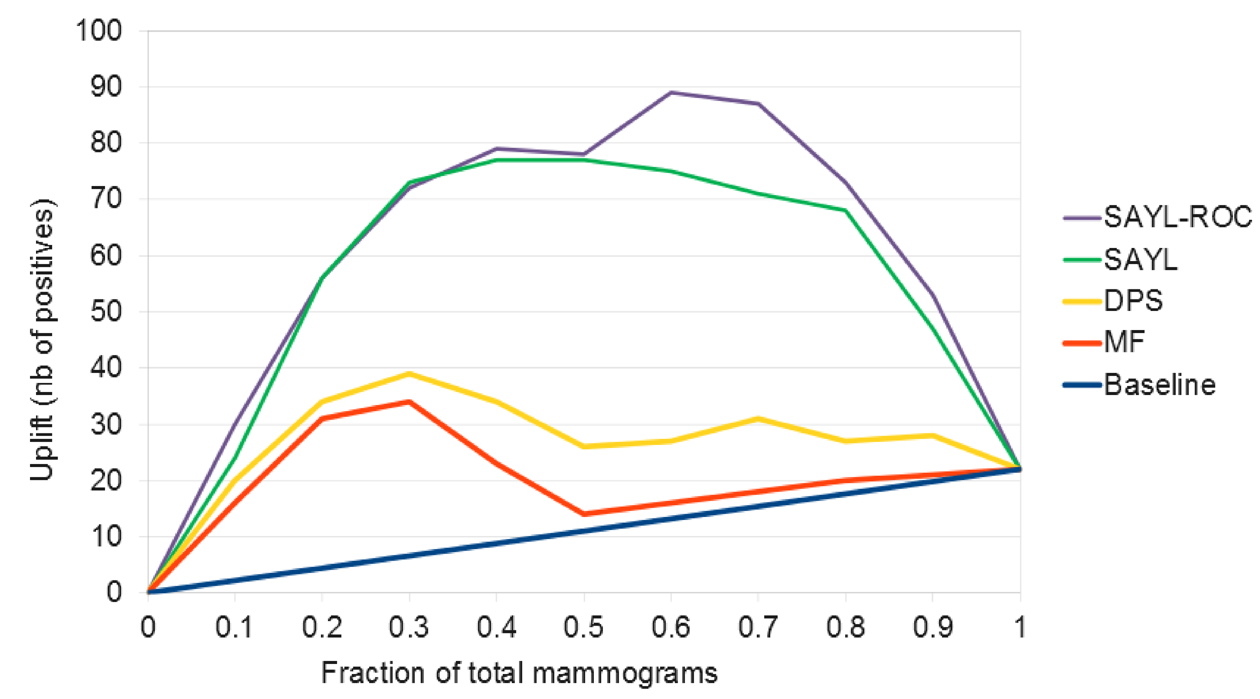
Acknowledgements

We thank NIH grant R01-CA165229, the Carbone Cancer Center, and NCI grant P30CA014520 for support.

Results

We test SAYL-ROC on a breast cancer mammography dataset. Our subject and control sets are respectively older and younger patients with confirmed breast cancer. Positive instances have in situ cancer, and negative instances have invasive cancer. The aim is to maximize the in situ cases' uplift.

The older cohort has 132 in situ and 401 invasive cases, while the younger one has 110 in situ and 264 invasive. The skews are $P_s = 132, \pi_s = \frac{132}{132+401}$ (older), and $P_c = 110, \pi_c = \frac{110}{110+264}$ (younger).



We use 10-fold cross-validation, making sure all records pertaining to the same patient are in the same fold. We run both SAYL and SAYL-ROC with a time limit of one hour per fold. For each cross-validated run, we use 4 training, 5 tuning folds and 1 testing fold. For each fold, we used the best combination of parameters according to a 9-fold internal cross-validation using 4 training, 4 tuning and 1 testing folds. We try two search modes, vary minpos between 7 and 13 (respectively 5% and 10% of older in situ examples), and set θ to 1%, 5% and 10%. We evaluate the final SAYL and SAYL-ROC models using their final uplift curves, concatenated from the results of each testing set. The table compares SAYL-ROC and SAYL to the previous ILP-based methods, Differential Prediction Search (DPS) and Model Filtering (MF), using minpos of 13. Their uplift areas are compared using the paired Mann-Whitney test at 95% confidence.

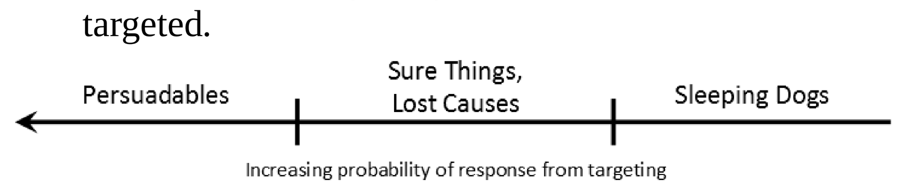
Algorithm	AUU	AUL_s	AUL_c	Rules Avg.	p-value
SAYL-ROC	62.99	95.64	32.65	24.7	0.4316
SAYL	58.10	97.24	39.15	9.3	-
DPS	27.83	101.01	73.17	37.1	0.0020 *
MF	20.90	100.89	80.99	19.9	0.0020 *
Baseline	11.00	66.00	55.00	-	0.0020 *

Uplift Modeling

Uplift modeling is a differential prediction technique that comes from marketing. In marketing, customers are broken into four categories:

Persuadables
  Customers who will respond only when targeted.
Sure Things
  Customers who will respond even when not targeted.
Lost Causes
  Customers who will not respond, regardless of whether they are targeted or not.
Sleeping Dogs
  Customers who will not respond as a result of being targeted.

Only *Persuadables* and *Sleeping Dogs* have any effect on the value produced by a marketing action, and, ideally, only *Persuadables* would be targeted.



Unfortunately, however, the group to which a customer belongs is unknown. Only the customer response and whether they were targeted can be observed experimentally.

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

To differentiate *Persuadables* from *Sure Things* and *Sleeping Dogs*, models are trained to better predict response in the targeted group than the control group. The assumption is that such a model captures the characteristics that are more specific to the *Persuadables*.

We wish to use this technique to capture the characteristics that are specific to older patients with in situ breast cancer.

The difference in performance between the targeted and control groups is often measured using uplift, the difference in lift. Lift is not a common metric though and is perhaps less approachable or understandable to those outside of the marketing domain.

Lift and ROC AUC

There is a strong connection between AUL and AUC. Let $\pi = \frac{P}{P+N}$ be the positive class skew, then:

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

Uplift modeling aims at optimizing uplift, the difference in lift over two sets.

$$AUU = AUL_s - AUL_c = \Delta(AUL)$$

It constructs a new classifier such that:

$$\Delta(AUL^*) > \Delta(AUL)$$

Expanding this, we get:

$$AUL_s^* - AUL_c^* > AUL_s - AUL_c$$

Which is equivalent to:

$$\frac{AUC_s^* - AUC_s}{AUC_c^* - AUC_c} > \frac{P_c \times (1-\pi_c)}{P_s \times (1-\pi_s)}$$

In a balanced dataset, we have $\pi_c = \pi_s = \frac{1}{2}$ and $P_c = P_s$, so we have $\frac{P_c \times (1-\pi_c)}{P_s \times (1-\pi_s)} = 1$. Thus, if the subject and control sets have the same numbers and skew:

$$\Delta(AUL^*) > \Delta(AUL)$$
$$\rightarrow \Delta(AUC^*) > \Delta(AUC)$$

Otherwise, no such guarantee can be made.

Conclusions

SAYL and SAYL-ROC significantly outperform previous methods, but there is no significant difference between the two. Even though SAYL- ROC is optimizing for $\Delta(AUC)$ during its training phase, it performs just as well as SAYL, which optimizes for $\Delta(AUL)$.

This result suggests that, on a moderately subject/control skewed data, $\Delta(AUC)$ can indeed be used for uplift modeling. ROC is more frequently used than lift, and may be more intuitive in many domains. Nevertheless, more experiments are needed to establish ROC-based uplift performance. We plan on measuring $\Delta(AUL)$ vs. $\Delta(AUC)$ for various skews.

SAYL-ROC produces as many rules as other ILP-based methods, more than twice that of SAYL. It is easy to interpret the final theory of other ILP methods as all of the rules are independent. SAYL and SAYL-ROC rules, however, are conditioned on each other as nodes in a Bayesian network, decreasing rule interpretability. At an average of 9.3 rules, a SAYL model is likely more interpretable, whereas at 24.7, SAYL-ROC sacrifices interpretability.

In conclusion, SAYL-ROC exhibits a similar performance to SAYL on our data, suggesting that ROC can be used for uplift modeling. SAYL-ROC returns larger models, reducing interpretability. More experiments are needed to test ROC-based uplift over different subject/control skews.

Selected References

Davis, J., Burnside, E., de Castro Dutra, I., Page, D., Santos Costa, V.: An integrated approach to learning Bayesian Networks of rules. In: Proceedings of the 16th European Conference on Machine Learning. pp. 84–95. Porto, Portugal (2005).

Hansotia, B., Rukstales, B.: Incremental value modeling. Journal of Interactive Marketing 16(3), 35–46 (2002).

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 (ECML-PKDD). Prague (2013).

Nassif, H., Santos Costa, V., Burnside, E., Page, D.: Relational differential prediction. In: European Conference on Machine Learning (ECML-PKDD). pp. 617–632. Bristol, UK (2012).

Radcliffe, N. and Simpson. R.: Identifying who can be saved and who will be driven away by retention activity. Journal of Telecommunications Management 1(2), 168–176 (2008).

Radcliffe, N., Surry, P.: Differential response analysis: Modeling true response by isolating the effect of a single action. In: Credit Scoring and Credit Control VI. Edinburgh, Scotland (1999).

Radcliffe, N., Surry, P.: Real-world uplift modelling with significance-based uplift trees. White Paper TR-2011-1, Stochastic Solutions (2011).

Rzepakowski, P., Jaroszewicz, S.: Decision trees for uplift modeling with single and multiple treatments. Knowledge and Information Systems 32, 303–327 (2012).

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