

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/220765228>

# Decision Trees for Uplift Modeling

Conference Paper · December 2010

DOI: 10.1109/ICDM.2010.62 · Source: DBLP

---

CITATIONS

42

---

READS

3,636

2 authors:



[Piotr Rzepakowski](#)

7 PUBLICATIONS 237 CITATIONS

SEE PROFILE



[Szymon Jaroszewicz](#)

Polish Academy of Sciences

62 PUBLICATIONS 1,159 CITATIONS

SEE PROFILE

# Decision trees for uplift modeling

Piotr Rzepakowski

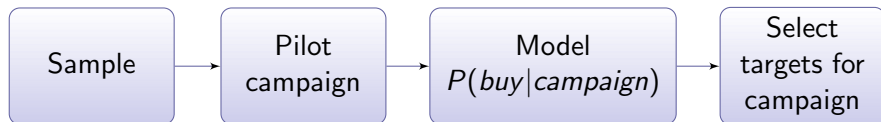
National Institute of Telecommunications  
Warsaw, Poland  
Warsaw University of Technology  
Warsaw, Poland

Szymon Jaroszewicz

National Institute of Telecommunications  
Warsaw, Poland  
Polish Academy of Sciences  
Warsaw, Poland

ICDM 2010

# Marketing campaign example



# Main idea of uplift modeling

We can divide objects into four groups

- 1 Responded **because** of the action
- 2 Responded **regardless** of whether the action is taken (**unnecessary costs**)
- 3 Did not respond and the action had **no impact** (**unnecessary costs**)
- 4 Did not respond **because** the action had a **negative impact** (e.g. customer got annoyed by the campaign, may even churn)

# Traditional classification vs. uplift modeling

Traditional models predict the conditional probability

$$P(\text{response}|\text{treatment})$$

# Traditional classification vs. uplift modeling

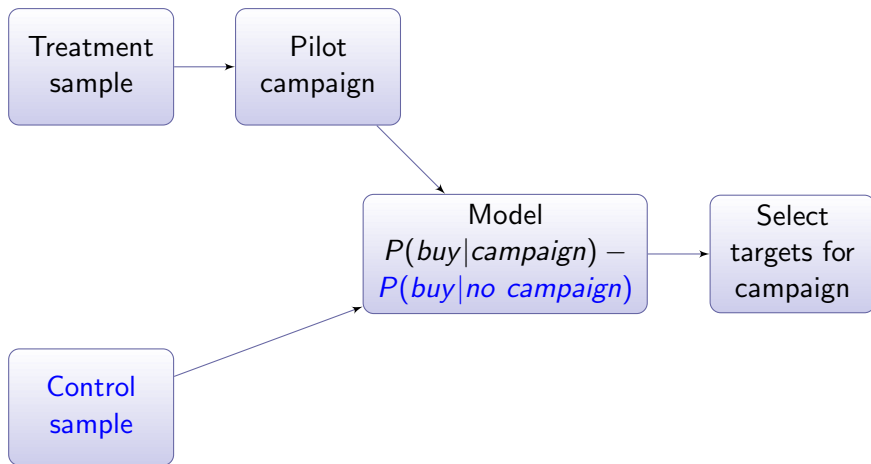
Traditional models predict the conditional probability

$$P(\text{response}|\text{treatment})$$

Uplift models predict change in behaviour resulting from the action

$$P(\text{response}|\text{treatment}) - P(\text{response}|\text{no treatment})$$

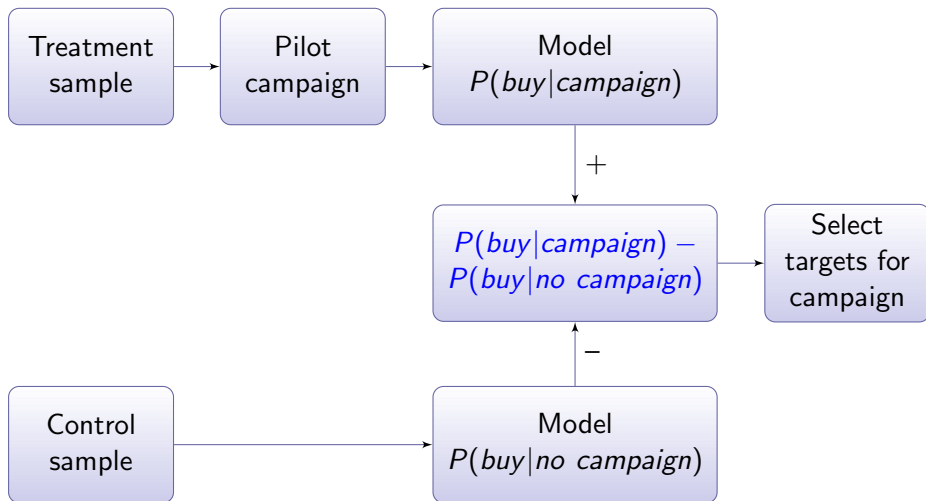
# Marketing campaign example (uplift modeling approach)



- Literature
  - Surprisingly little attention in literature
  - Business whitepapers offering vague descriptions of algorithms used
- Two general approaches
  - Subtraction of two models
  - Modification of model learning algorithms

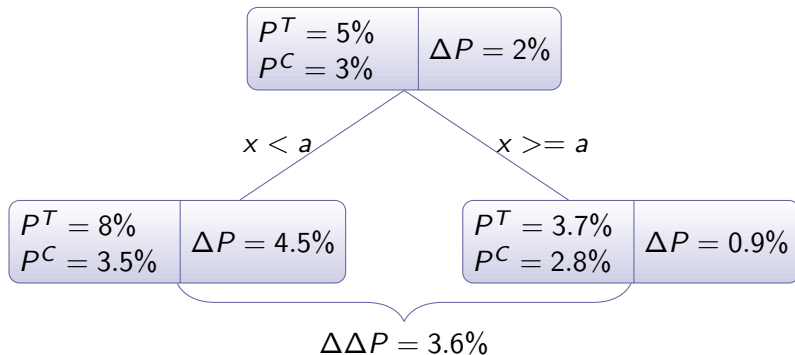


# Subtraction of two models



# Current approaches to uplift decision trees

- Create splits using difference of probabilities ( $\Delta\Delta P$ )



- Pruning not used (or not described)
- Work only for two class problems and binary splits

# Our approach to uplift decision trees

- Splitting criteria based on **Information Theory**
- **Pruning** strategy designed for uplift modeling
- **Multiclass** problems and multiway splits possible
- If the **control group is empty**, the criterion should reduce to one of classical splitting criteria used for decision tree learning

# Kullback-Leibler divergence

- Measure difference between treatment and control groups using KL divergence

$$KL\left(P^T(Class) : P^C(Class)\right) = \sum_{y \in \text{Dom}(Class)} P^T(y) \log \frac{P^T(y)}{P^C(y)}$$

# Kullback-Leibler divergence

- Measure difference between treatment and control groups using KL divergence

$$KL\left(P^T(Class) : P^C(Class)\right) = \sum_{y \in \text{Dom}(Class)} P^T(y) \log \frac{P^T(y)}{P^C(y)}$$

- Need KL-divergence conditional on a given test

$$\begin{aligned} & KL(P^T(Class) : P^C(Class) | Test) \\ &= \sum_{a \in \text{Dom}(Test)} \frac{N^T(a) + N^C(a)}{N^T + N^C} KL\left(P^T(Class|a) : P^C(Class|a)\right) \end{aligned}$$

Measures how much the two groups differ given a test's outcome

# Final splitting criterion

$$KL_{gain}(Test) = KL(P^T(Class) : P^C(Class) | Test) - KL(P^T(Class) : P^C(Class))$$

- Measures the *increase* in difference between treatment and control groups from splitting based on *Test*
- If the control group is empty,  $KL_{gain}$  reduces to entropy gain

# Final splitting criterion

$$KL_{gain}(Test) = KL(P^T(Class) : P^C(Class) | Test) - KL(P^T(Class) : P^C(Class))$$

- Measures the *increase* in difference between treatment and control groups from splitting based on *Test*
- If the control group is empty,  $KL_{gain}$  reduces to entropy gain

$$KL_{ratio} = \frac{KL_{gain}(Test)}{KL_{value}(Test)}$$

- Tests with **large number of values** are punished
- Tests which split the control and treatment groups in **different proportions** are punished
- Postulates are satisfied

# Splitting criterion based on squared Euclidean distance

$$Euclid \left( P^T(Class) : P^C(Class) \right) = \sum_{y \in \text{Dom}(Class)} \left( P^T(y) - P^C(y) \right)^2$$

- $Euclid_{gain}$ ,  $Euclid_{ratio}$  analogous to  $KL$
- Better statistical properties (values are bounded)
- Symmetry



# Pruning procedure (maximum class probability difference)

- Definitions

$$\text{Diff}(\text{Class}, \text{node}) = P^T(\text{Class}|\text{node}) - P^C(\text{Class}|\text{node})$$

- Maximum class probability difference (MD)

$$MD(\text{node}) = \max_{\text{Class}} |\text{Diff}(\text{Class}|\text{node})|$$

$$\text{sign}(\text{node}) = \text{sgn}(\text{Diff}(\text{Class}^*, \text{node}))$$

# Pruning procedure (maximum class probability difference)

- Definitions

$$Diff(Class, node) = P^T(Class|node) - P^C(Class|node)$$

- Maximum class probability difference (MD)

$$MD(node) = \max_{Class} |Diff(Class|node)|$$

$$sign(node) = \text{sgn}(Diff(Class^*, node))$$

- Use separate validation sets

- Bottom up procedure

- Keep subtree if

- On validation set:  $MD$  of the subtree is greater than if it was replaced with a leaf
- And** the sign of  $MD$  is the same in training and validation sets

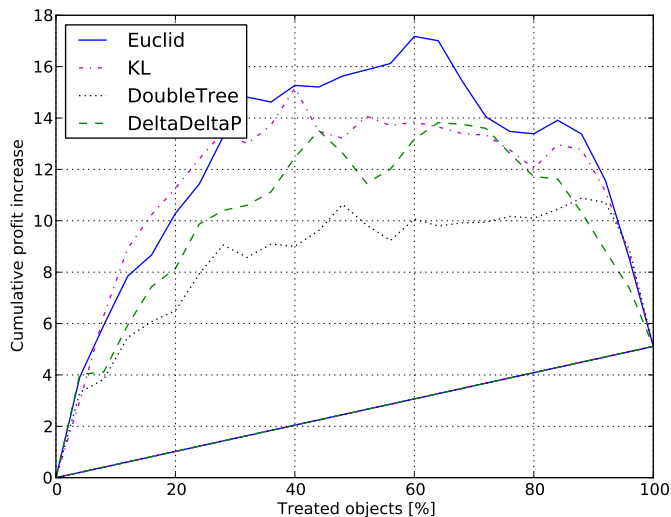
- Compared models

- 1 **Euclid** - uplift decision trees based on  $E_{ratio}$
- 2 **KL** - uplift decision trees based on  $KL_{ratio}$
- 3 **DeltaDeltaP** - based on the  $\Delta\Delta P$  criterion
- 4 **DoubleTree** - separate decision trees for the treatment and control groups

# Method of evaluating uplift classifiers

- Control and treatment datasets are scored using the same model
- Compute **lift curves** on both datasets
- **Uplift curve** = lift curve on treatment data – lift curve on control data
- Measure model's performance based on
  - Area under the uplift curve (AUUC)
  - Height of the uplift curve at the 40th percentile

# The uplift curve for the splice dataset



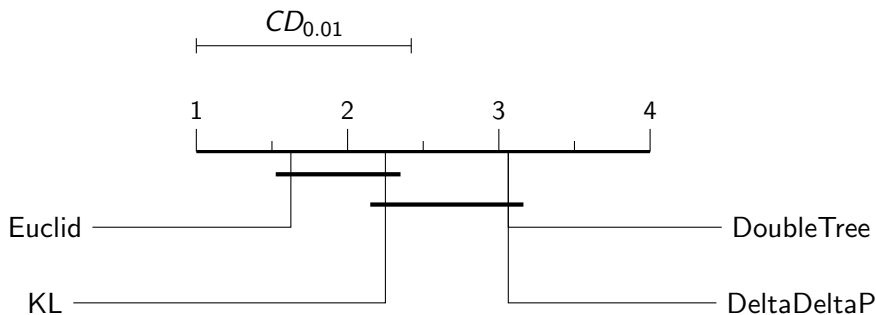
- Lack publicly available data to test uplift models
- Datasets from UCI repository were split into treatment and control groups based on one attribute
- Procedure of choosing the splitting attribute:
  - If an action was present it was picked (e.g. hepatitis data)
  - Otherwise pick the first attribute which gives a reasonably balanced split

# Methodology of model comparison

- 1 Models are evaluated using  $2 \times 5$  **crossvalidation**
- 2 Models are compared by ranking on all datasets
- 3 Check if there are differences in model performance using **Friedman's test**, a nonparametric analogue of ANOVA
- 4 If the test shows significant differences, a post-hoc **Nemenyi test** is used to assess which of the models are significantly different

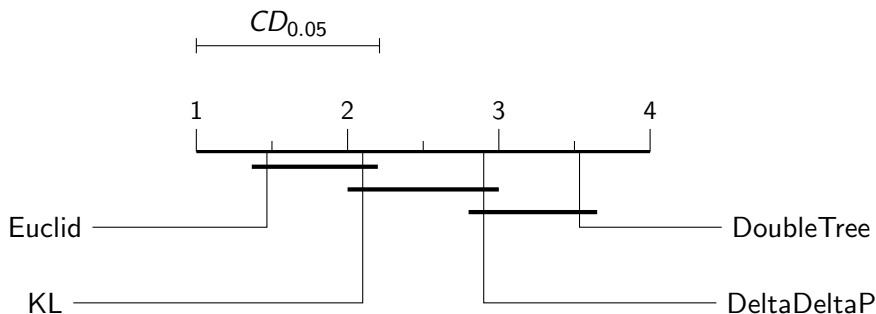
# Results for Area Under Uplift Curve

## Nemenyi test at $p = 0.01$





# Results for the height of the curve at the 40th percentile Nemenyi test at $p = 0.05$



- Method for decision tree construction for uplift modeling in the style of modern decision tree learning
  - Information Theory based splitting
  - Dedicated pruning strategy
- Two splitting criteria (KL and Euclidian distance)
- Reduce to standard decision trees if control data absent
- The new method significantly outperforms previous approaches to uplift modeling
- Other applications e.g. medicine