

# Modelowanie różnicowe

Szymon Jaroszewicz

Instytut Podstaw Informatyki PAN

## Streszczenie

W praktyce, celem modelowania statystycznego i uczenia maszynowego jest wytypowanie obiektów (klientów, pacjentów) których należy poddać jakiemuś działaniu, np. terapii czy akcji marketingowej. Stosowane powszechnie modele nie są jednak do tego celu dostosowane, gdyż nie biorą pod uwagę wyniku w sytuacji gdy dany obiekt nie zostanie działaniu poddany. Dla przykładu, pacjent mógłby wyzdrowieć również bez zastosowania terapii, jej skutki uboczne mogły nawet być gorsze od samej choroby. Jednym ze sposobów rozwiązania tego problemu jest uwzględnienie w trakcie budowy modelu grupy kontrolnej nie poddanej danemu działaniu. Modelowanie różnicowe zajmuje się predykcją różnicy między prawdopodobieństwami sukcesu po poddaniu danego obiektu działaniu i rezygnacji z niego. Pozwala to na przyczynowe modelowanie skutków działań na poziomie pojedynczych obiektów.

W ramach wykładu przedstawiony zostanie model przyczynowy Rubina, który stanowi teoretyczne ramy podejścia. Następnie omówione zostaną metody typowo statystyczne: podejście dwumodelowe, g-formuła i g-estymacja. W drugiej części wykładu przedstawione zostaną podejścia pochodzące z dziedziny uczenia maszynowego, takie jak drzewa decyzyjne, metody oparte o transformację zmiennej celu, zespoły modeli: bagging, boosting i lasy losowe. Przedstawione zostaną również zastosowania do danych przeżycia i metody oceny modeli różnicowych.