

Łukasz Zaniewicz

Support Vector Machines
for Uplift Modeling

PhD dissertation

Supervisor: dr hab. inż. Szymon Jaroszewicz

Institute of Computer Science
Polish Academy of Sciences
Warsaw, January 2018

Research of the author was supported by Project 'Information technologies: research and their interdisciplinary applications' POKL.04.01.01-00-051/10-00.

Abstract

Łukasz Zaniewicz, *Support Vector Machines for Uplift Modeling* Doctoral dissertation supervised by dr hab. inż. Szymon Jaroszewicz, Institute of Computer Science, Polish Academy of Sciences, Warsaw 2018.

Uplift modeling is a branch of Machine Learning which aims to predict not the class itself, but the difference between the class variable behavior in two groups: treatment and control. Objects in the treatment group have been subjected to some action, while objects in the control group have not. By including the control group it is possible to build a model which predicts the *causal* effect of the action for a given individual. As a consequence, uplift modeling is directly applicable in fields where presence of the control group is common in market practice. This is always the case for randomized controlled clinical trials and is becoming a standard in marketing campaigns where it has been realized that, at least in some situations, the true effect of an examined action can only be measured against a background of a control group.

This dissertation presents application of the Support Vector Machines designed specifically for uplift modeling. The SVM optimization task has been reformulated to explicitly model the difference in class behavior between two datasets. The model predicts whether a given object will have a positive, neutral or negative response to a given action, and by tuning a parameter of the model the analyst is able to influence the relative proportion of neutral predictions and thus the conservativeness of the model. Moreover, this work extends also the L_p -SVMs to the case of uplift modeling and demonstrates that this allows for a more stable selection of the size of negative, neutral and positive groups. Furthermore, efficient quadratic and convex optimization methods are presented for efficiently solving the two related optimization tasks. Experiments demonstrate that the proposed methods compare favorably with other uplift modeling approaches.

This dissertation discusses also the issue of nonrandom assignment to treatment and control groups. In general, uplift modeling is best applied to training sets obtained from randomized controlled experiments, but such experiments are not always possible, in which case treatment

assignment is biased. To handle such situations we proposed a modification of the Uplift Support Vector Machines which are less sensitive to such a bias. This is achieved by including in the model formulation an additional term which penalizes models which score treatment and control groups differently. We call the technique Székely regularization since it is based on the energy distance proposed by Székely and Rizzo. Optimization algorithm based on stochastic gradient descent techniques has been developed for this problem. Further, this work demonstrates experimentally that the proposed regularization term does indeed produce uplift models which are less sensitive to biased treatment assignment.

Streszczenie

Łukasz Zaniewicz, *Maszyny wektorów wspierających w modelowaniu różnicowym*. Rozprawa doktorska przygotowana pod kierunkiem dr. hab. inż. Szymona Jaroszewicza, Instytut Podstaw Informatyki Polskiej Akademii Nauk, Warszawa 2018.

Modelowanie różnicowe jest dziedziną uczenia maszynowego, której celem nie jest przewidywanie zmiennej celu, lecz różnic w zachowaniu zmiennej celu między dwoma grupami: eksperymentalną i kontrolną. Obiekty w grupie eksperymentalnej zostały poddane pewnemu działaniu, podczas gdy obiekty w grupie kontrolnej nie. Włączając grupę kontrolną, możliwe staje się zbudowanie modelu przewidującego *przyczynowy* efekt danego działania na poziomie konkretnego obiektu. W konsekwencji modelowanie różnicowe znajduje bezpośrednie zastosowanie w dziedzinach, w których stosowanie grup kontrolnych jest powszechne. Jest tak w przypadku randomizowanych prób klinicznych, staje się też standardem w marketingu bezpośrednim, gdzie zauważono, że, przynajmniej w niektórych sytuacjach, prawdziwą korzyść z badanego działania można zmierzyć tylko biorąc pod uwagę grupę kontrolną.

Niniejsza praca przedstawia maszyny wektorów wspierających zaadaptowane do problemu modelowania różnicowego. Zadanie optymalizacyjne maszyn klasyfikacyjnych zostało sformułowane, tak aby możliwe było modelowanie różnic w zachowaniu zmiennej celu między dwoma zbiorami danych. Model przewiduje, czy wpływ badanego działania na dany obiekt będzie pozytywny, neutralny czy negatywny. Poprzez odpowiedni dobór parametrów modelu, analityk ma bezpośredni wpływ na względny udział prognoz neutralnych. Dalej, w pracy dostosowano do problemu modelowania różnicowego także model L_p -SVMs, wykorzystujący zmodyfikowaną funkcję straty. Pokazano, że to pozwala na bardziej stabilny wybór proporcji przewidywań negatywnego, neutralnego i pozytywnego wpływu działania. Przedstawiono również wydajne metody optymalizacji kwadratowej i wypukłej służące do rozwiązywania problemów optymalizacyjnych związanych z proponowanymi modelami. Eksperymentalnie wykazano, że proponowane algorytmy są konkurencyjne wobec innych metod modelowania różnicowego.

W rozprawie omówiony został również problem nielosowego przypisania obiektów do grupy eksperymentalnej i kontrolnej. Zasadniczo, modelowanie różnicowe najlepiej jest stosować w przypadku przypisania całkowicie losowego, ale takie eksperymenty nie zawsze są możliwe. W praktyce, przydział do grupy kontrolnej często następuje na podstawie decyzji dokonywanych przez człowieka, jest zatem obciążony. Aby możliwe było zastosowanie modelowania różnicowego w takim scenariuszu, w pracy przedstawiono modyfikację proponowanego modelu, która jest mniej wrażliwa na obciążony podział na grupy. Modyfikacja polega na dodaniu do sformułowania modelu dodatkowego wyrazu, który karze modele, których predykcje w grupie kontrolnej i eksperymentalnej różnią się. Technikę tę nazywano regularyzacją Székely’ego, ponieważ opiera się ona na odległości energetycznej zaproponowanej przez Székely’ego i Rizzo. Przedstawiono również algorytm optymalizacji dla tego typu modeli oparty o techniki stochastycznego spadku gradientu. Ponadto, w pracy wykazano eksperymentalnie, że proponowany dodatkowy wyraz regularyzacyjny istotnie tworzy modele różnicowe, które są mniej wrażliwe na nielosowe przypisanie obiektów do grup eksperymentalnej i kontrolnej.