

Learning to rank w modelowaniu różnicowym i medycynie spersonalizowanej

dr hab. inż. Szymon Jaroszewicz

Instytut Podstaw Informatyki PAN
s.jaroszewicz@ipipan.waw.pl

Celem uczenia maszynowego jest często wybieranie celów dla określonego działania takiego jak terapia medyczna lub kampania promocyjna. Typowe przykłady to medycyna spersonalizowana i kampanie marketingu bezpośredniego. Niestety klasyczne metody uczenia maszynowego nie nadają się do tego zadania, ponieważ nie są w stanie określić jego rzeczywistego efektu. Powodem jest to, że się nie biorą pod uwagę, wyniku w przypadku gdyby działanie nie zostało podjęte. Rozwiązaniem tego problemu jest zastosowanie drugiego zbioru danych treningowych zawierającego grupę kontrolną, która nie została danemu działaniu poddana. Celem jest budowa modelu, który przewiduje różnicę między zachowaniami osób poddanych i nie poddanych działaniu: czyli realną korzyść z jego podjęcia. Takie podejścia znane są jako **modelowanie różnicowe** (ang. *uplift modeling*) lub *heterogeneous treatment effect estimation*.

Większość klasycznych modeli uczenia maszynowego została dostosowana do problemu modelowania różnicowego [2, 3], istnieje jednak szereg otwartych obszarów badawczych. Jednym z nich jest tzw. *learning to rank*. Modele uczenia maszynowego są zazwyczaj używane do szeregowania potencjalnych celów (pacjenci, klienci) według przewidywanego prawdopodobieństwa sukcesu. *Learning to rank* pozwala na bezpośrednie nauczenie się prawidłowej kolejności zamiast uczenia się funkcji skorującej, na podstawie której przypadki będą sortowane [1]. Ponieważ większość metod modelowania różnicowego służy do wyboru najlepszych celów dla danego działania, dostosowanie metod *learning to rank* do przypadku modelowania różnicowego jest ważne z punktu widzenia zastosowań.

Inne możliwe tematy badawcze obejmują opracowanie teorii uczenia się dla modelowania różnicowego lub opracowanie bardziej zaawansowanych metod opartych o drzewa decyzyjne i ich komitety.

Literatura

- [1] Hang Li. A short introduction to learning to rank. *IEICE Transactions on Information and Systems*, E94.D(10):1854–1862, 2011.

- [2] Krzysztof Rudaś and Szymon Jaroszewicz. Linear regression for uplift modeling. *Data Mining and Knowledge Discovery*, 32(5):1275–1305, Sep 2018.
- [3] Ł. Zaniewicz and S. Jaroszewicz. Lp-support vector machines for uplift modeling. *Knowledge and Information Systems*, 53(1):269–296, Oct 2017.