

Predykcja i selekcja zmiennych w klasyfikacji wieloetykietowej

Paweł Teisseyre

W klasycznym problemie klasyfikacji modelujemy zależność między zmienną odpowiedzi (najczęściej binarną) a zmiennymi objaśniającymi. W klasyfikacji wieloetykietowej rozważamy wiele binarnych zmiennych odpowiedzi jednocześnie. W ostatnich latach klasyfikacja wieloetykietowa wzbudziła bardzo duże zainteresowanie [1]. Metody klasyfikacji wieloetykietowej są stosowane w wielu dziedzinach, takich jak automatyczna kategoryzacja tekstów, rozpoznawanie obrazów, modelowanie wielozachorowalności i wiele innych. Istotnym wyzwaniem w klasyfikacji wieloetykietowej jest zbudowanie modelu o dużej mocy predykcyjnej oraz selekcja istotnych zmiennych. W referacie przedstawię nową metodę [2], opartą o wykorzystanie łańcuchów klasyfikatorów oraz regresji logistycznej z regularyzacją. Podstawową zaletą przedstawionej metody jest to że selekcja istotnych zmiennych jest wykonywana podczas dopasowania modelu. Przedstawię wyniki teoretyczne dotyczące stabilności i oszacowania błędu generalizacji oraz wyniki eksperymentów na rzeczywistych danych.

P. Teisseyre, INSTYTUT PODSTAW INFORMATYKI PAN
Adres e-mail: teisseyre@ipipan.waw.pl

Literatura

- [1] S. Ventura E. Gibaja, *A tutorial on multilabel learning*, ACM Comput. Surv. **47** (2015), 1–38.
- [2] P. Teisseyre, *CCnet: joint multi-label classification and feature selection using classifier chains and elastic net regularization*, Neurocomputing **235** (2017), 98–111.