

Model łańcuchowy w klasyfikacji wieloetykietowej

Paweł Teisseyre

Instytut Podstaw Informatyki

Streszczenie

Klasyfikacja wieloetykietowa (ang. multi-label classification) odnosi się do sytuacji w której każdy z obserwowanych obiektów jest opisany wektorem zmiennych objaśniających oraz wektorem odpowiedzi (najczęściej binarnych) zwanych etykietami. Na podstawie danych uczących buduje się model który umożliwi przewidywanie etykiet dla nowych obiektów. W ostatnich kilku latach problem klasyfikacji wieloetykietowej wzbudził bardzo duże zainteresowanie. Metody klasyfikacji wieloetykietowej znajdują zastosowanie w takich dziedzinach jak: kategoryzacja tekstów, automatyczna anotacja obrazów, przewidywanie współwystępowania chorób i wiele innych. W referacie przedstawię model łańcuchowy, w którym prawdopodobieństwo wystąpienia danej etykiety jest opisane logistyczną funkcją odpowiedzi, zależną od wektora zmiennych objaśniających i etykiet poprzedzających daną etykietę w łańcuchu. Nietrywialnym zadaniem, związanym z przedstawionym podejściem, jest wybór optymalnej kolejności budowy modeli. Zaproponowany algorytm wyznaczania kolejności opiera się na wykorzystaniu miar specyfikacji modelu logistycznego. Przedstawione zostaną również wyniki eksperymentów symulacyjnych przeprowadzonych na rzeczywistych zbiorach danych.

Literatura

- Teisseyre, P. (2014). Asymptotic consistency and order specification for logistic classifier chains in multi-label learning. *w recenzji*.
- Read, et. al (2009). Classifier chains for multi-label classification. *Proceedings of the 20th European Conference on Machine Learning*, 254-269.