

# Klasyfikacja wieloetykietowa z niepełną obserwowalnością etykiet

**Paweł Teisseyre**

Instytut Podstaw Informatyki PAN oraz Wydział  
Matematyki i Nauk Informacyjnych PW

Przedstawię problem klasyfikacji wieloetykietowej z niepełną obserwowalnością etykiet. W tradycyjnym problemie klasyfikacji wieloetykietowej zakłada się że zbiór uczący zawiera poprawnie przypisane etykiety dla wszystkich obserwacji. To założenie może być w wielu sytuacjach niespełnione, np. w przypadku prognozowania wielozachorowalności (współwystępowania wielu chorób u jednego pacjenta) może się zdarzyć że pewne choroby nie zostały zdiagnozowane i nie ma wówczas pewności że występują lub nie u danego pacjenta. W rozważanym schemacie, przyjmujemy że jeśli etykieta została przypisana to na pewno jest związana z daną obserwacją, zaś brak przypisania może oznaczać zarówno to że jest ona związana z daną obserwacją jak i to że nie jest z nią związana. Omówię podstawowe własności teoretyczne takiego schematu. Pokażę w jaki sposób zmodyfikować łańcuchy klasyfikatorów (będące najpopularniejszym algorytmem w klasyfikacji wieloetykietowej) do powyższego schematu.

## **Literatura**

- [1] J. Bekker, J. Davis (2020), *Learning from positive and unlabeled data: a survey*, Machine Learning 109 (4), 719-760
- [2] P. Teisseyre (2020), *Classifier chains for positive unlabelled multi-label learning*, under review