

Marcin Czajkowski¹, Marek Krętowski¹

AN EXTENSION OF TSP-FAMILY ALGORITHMS FOR MICROARRAY CLASSIFICATION

Abstract: Classification of microarray data and generation of simple and efficient decision rules may be successfully performed with Top Scoring Pair algorithms. *TSP*-family methods are based on pairwise comparisons of gene expression values. This paper presents a new method, referred as *Linked TSP* that extends previous approaches *k – TSP* and *Weight k – TSP* algorithms by linking top pairwise mRNA comparisons of gene expressions in different classes. Opposite to existing *TSP*-family classifiers, the proposed approach creates decision rules involving single genes that most frequently appeared in top scoring pairs. Motivation of this paper is to improve classification accuracy results and to extract simple, readily interpretable rules providing biological insight as to how classification is performed. Experimental validation was performed on several human microarray datasets and obtained results are promising.

Keywords: pairwise classification, decision rules, microarray, gene expression

ROZSZERZENIE METOD Z RODZINY TSP W KLASYFIKACJI MIKROMACIERZY DNA

Streszczenie Klasyfikacja danych mikromacierzowych a także późniejsza interpretacja reguł decyzyjnych może być skutecznie przeprowadzona za pomocą metod z rodziny Top Scoring Pair, polegających na analizie par genów o przeciwstawnych poziomach ekspresji w różnych klasach. W poniższym artykule zaprezentowano nową metodę: *Linked TSP*, która rozszerza działanie klasyfikatorów *k – TSP* i *Weight k – TSP*. W przeciwieństwie do algorytmów z rodziny *TSP* proponowane rozwiązanie tworzy reguły decyzyjne zbudowane z pojedynczych genów co znacznie ułatwia ich późniejszą interpretację medyczną. W algorytmie wykorzystywane są pary genów uzyskane z algorytmów *TSP* z których następnie, wybierane są pojedyncze, najczęściej powtarzające się geny. Testy algorytmu *Linked TSP* przeprowadzone zostały na rzeczywistych zbiorach danych pacjentów a uzyskane wyniki są obiecujące.

¹ Faculty of Computer Science, Białystok Technical University, Poland

Słowa kluczowe: klasyfikacja par genów zależnych, analiza mikromacierzy, reguły decyzyjne, ekspresja genów

Artykuł zrealizowano w ramach pracy badawczej W/WI/5/08.