

Autor rozprawy doktorskiej: mgr inż. Paweł Foszner

Tytuł rozprawy doktorskiej w języku polskim:

Bi-klasteryzacja – algorytmy i ich zastosowania

Tytuł rozprawy doktorskiej w języku angielskim:

Bi-clustering – algorithms and applications

Promotor rozprawy doktorskiej: prof. dr hab. inż. Andrzej Polański

Jednostka prowadząca przewód doktorski:

Politechnika Śląska, Wydział Automatyki, Elektroniki i Informatyki

Słowa kluczowe:

bi-klasteryzacja, dwu-grupowanie, klasteryzacja, klasyfikacja

Streszczenie rozprawy doktorskiej w języku polskim:

Celem pracy było opracowanie uniwersalnego podejścia do analiz bi-klasteryzacji oraz uodpornienie metody na strukturę posiadanych danych. W tym celu stworzono zbiór danych syntetycznych który pokrywał niemal wszystkie istotne warianty danych. Uzyskane na ich podstawie wyniki wykazały że podejście zaproponowane w rozprawie jest wyraźnie lepsze od dostępnych metod lub nie gorsze niż 3 najlepsze dla zadanych danych algorytmy. Miarą jakości dla danych syntetycznych była średnia arytmetyczna z miary określającej pokrycie uzyskanych bi-klastrów znalezionych w zbiorze bi-klastrów oczekiwanych oraz miary określającej pokrycie bi-klastrów oczekiwanych w zbiorze bi-klastrów znalezionych. Zaproponowane metody wykazały również że mogą poprawiać wyniki dla danych rzeczywistych. W tym celu przeanalizowano dwa zupełnie różne zbiory danych dostępne w literaturze. Wykazano że podejście znacząco poprawia jakość otrzymywanych bi-klastrów.

Streszczenie rozprawy doktorskiej w języku angielskim:

The aim of this thesis was to develop a universal approach to the analysis of bi-clustering and method that is resistant on the structure of the data. For this purpose, the synthetic dataset that covered almost all relevant data variants was created. Obtained on their basis results showed that the approach proposed in the dissertation is clearly better than the available methods or no worse than the three best algorithms (for this specific data). A measure of the quality of the synthetic data was the arithmetic mean of the measure defining the coverage obtained bi-clusters in a set of expected bi-clusters and measures of determining the coverage of expected bi-clusters in a set of found bi-clusters. In other words arithmetic mean of relevance and recovery. The proposed method has also shown that it can improve performance for real data. For this purpose, analyzed two completely different sets of data available in the literature. It has been shown that this approach significantly improves the quality of the bi-clusters.