

**Autor rozprawy doktorskiej:** mgr inż. Jakub Nalepa

**Tytuł rozprawy doktorskiej:**

Genetic and memetic algorithms for selection of training sets for support vector machines

**Promotor rozprawy doktorskiej:** dr hab. inż. Michał Kawoluk

**Jednostka prowadząca przewód doktorski:**

Politechnika Śląska, Wydział Automatyki, Elektroniki i Informatyki

**Słowa kluczowe:**

uczenie maszynowe, maszyna wektorów podpierających, dobór zbioru treningowego, algorytm ewolucyjny, algorytm genetyczny, algorytm memetyczny

**Streszczenie rozprawy doktorskiej w języku polskim:**

Maszyna wektorów podpierających (ang. *support vector machine*, SVM) jest klasyfikatorem nadzorowanym, który znalazł zastosowanie w rozwiązywaniu wielu zadań związanych z rozpoznawaniem wzorców. Trening klasyfikatora polega na wyznaczeniu hiperpłaszczyzny decyzyjnej separującej dane (tj. wektory cech) należące do dwóch klas w zbiorze treningowym.

W pracy rozważany jest problem klasyfikacji dwuklasowej. Położenie hiperpłaszczyzny separującej jest zdefiniowane przez pewien reprezentatywny podzbiór pełnego zbioru treningowego, nazywany wektorami podpierającymi. Hiperpłaszczyzna jest następnie wykorzystana do klasyfikacji nowych danych. Złożony proces treningu klasyfikatora staje się jego istotną wadą i utrudnia jego wykorzystanie w praktyce. Dodatkowo liczba wyznaczonych wektorów podpierających – która jest proporcjonalna do wielkości zbioru treningowego – wpływa na czas klasyfikacji. Oznacza to, że duża wielkość zbioru treningowego pośrednio wpływa na czas klasyfikacji.

W pracy przedstawiono pięć algorytmów ewolucyjnych do doboru zredukowanych zbiorów treningowych dla klasyfikatora SVM – algorytm genetyczny, adaptacyjny algorytm genetyczny, dynamicznie adaptacyjny algorytm genetyczny, algorytm memetyczny, oraz adaptacyjny algorytm memetyczny wykorzystujący dodatkową analizę geometrii danych w zbiorze treningowym. Zaproponowane algorytmy zostały wszechstronnie zbadane teoretycznie i eksperymentalnie. Wyniki badań dowiodły, że zredukowane zbiory treningowe otrzymane przy pomocy zaproponowanych algorytmów są lepsze niż zbiory wyekstrahowane przy pomocy przeanalizowanych metod znanych z literatury. Zaproponowane algorytmy ewolucyjne pozwalają nie tylko na zwiększenie skuteczności klasyfikacji, ale też na zmniejszenie liczby wektorów podpierających.

### **Streszczenie rozprawy doktorskiej w języku angielskim:**

Support vector machine (SVM) is a supervised classifier which has been applied for solving a wide range of pattern recognition problems. SVMs classify the incoming data based on the hyperplane defined by a subset of the training set vectors (called support vectors, SVs). The process of determining SVs (i.e., SVM training) is a constrained quadratic programming problem. Since real-life sets can very easily become enormously large in various domains, the time-consuming training is a major shortcoming of SVMs. The SVM response time is dependent on the number of SVs. Minimizing  $s$  while keeping the classification score high reduces the response time and makes SVMs suitable for real-time applications.

In this dissertation, five evolutionary algorithms for selection of training sets for support vector machines were proposed – the genetic algorithm, the adaptive genetic algorithm, the dynamically adaptive genetic algorithm, memetic algorithm, and the adaptive memetic algorithm enhanced with data geometry analysis of the training set. They were extensively analyzed theoretically and experimentally. The results showed that the reduced training sets retrieved using the proposed techniques are better than those elaborated using other state-of-the-art methods. The evolutionary methods allows for improving the SVM classification performance, and for decreasing the numbers of support vectors.