



Paweł Delimata¹, Zbigniew Suraj^{1,2}

(¹ Uniwersytet Rzeszowski, ² Wyższa Szkoła Informatyki i Zarządzania w Rzeszowie)

O metodzie k -NN z wstępnym przetwarzaniem

Celem referatu jest przedstawienie nowego modelu klasyfikacji danych z tzw. wstępnym przetwarzaniem. Model ten jest modyfikacją klasycznej metody k -najbliższych sąsiadów (metody k -NN)[1,2], w której zakłada się, że dla zadanego zbioru danych globalna metryka została zdefiniowana i wyszukiwanie najbliższych sąsiadów oparte jest na globalnej metryce. W pracy [3] zaproponowano model klasyfikacji danych oparty na metodzie k -NN z lokalną indukcją metryki. Na ogół, metoda ta w porównaniu z metodą klasyczną charakteryzuje się lepszym współczynnikiem jakości klasyfikacji nowych danych. Tym niemniej jednak dla dużych zbiorów danych jest ona mniej efektywna niż metoda klasyczna. O efektywności algorytmów klasyfikacji opartych na tych metodach decyduje nie tylko wybór metryki, ale także wielkość zbiorów danych wyjściowych przeznaczonych do analizy.

W referacie zaproponowane zostaną trzy metody wstępnego przetwarzania danych w celu zmniejszenia rozmiaru danych wyjściowych, nie tracąc przy tym istotnych informacji związanych z klasyfikacją nowych danych. Proponowane metody użyto w połączeniu z klasyczną metodą k -NN oraz metodą k -NN z lokalną indukcją metryki. Wyniki uzyskane w eksperymentach porównano z wynikami otrzymanymi dla tych samych danych, lecz bez wstępnego przetwarzania. Zredukowany zbiór danych umożliwił znaczne przyspieszenie działania algorytmów klasyfikujących (klasyfikatorów) opartych na rozważanych metodach k -NN przy zachowaniu wysokiego współczynnika jakości klasyfikacji.