

Wybrane zagadnienia klasyfikacji danych niezbalansowanych – bieżące trendy i wyzwania

prof. dr hab. inż. Michał Woźniak, Politechnika Wroclawska

Zbiorami niezbalansowanymi nazywamy takie zbiory, w których przykłady z niektórych klas występują zdecydowanie częściej (klasy większościowe), niż przykłady z pozostałych klas (klasy mniejszościowe). Tradycyjne metody klasyfikacji wzorców są dedykowane problemom, gdzie nie występują duże różnice w częstościach pojawiania się obiektów z różnych klas, jednakże w praktycznych zastosowaniach takie przypadki są dość częste. W przypadku problemu dwuklasowego związki pomiędzy klasami są łatwe do zdefiniowania, gdyż jedna z klas jest klasą mniejszościową, a druga klasą większościową. W przypadku zadania klasyfikacji wieloklasowej relacja pomiędzy klasami nie musi być już tak oczywista, gdyż jedna z klas może mieć charakter klasy większościowej w stosunku do niektórych klas, a w stosunku do innych może być klasą mniejszościową.

W trakcie wykładu zostaną przedstawione najważniejsze zagadnienia związane z problematyką klasyfikacji danych niezbalansowanych, zarówno dla klasyfikacji binarnej, jak i wieloklasowej, a także zostaną wskazane potencjalne nurty badawcze. Przedstawione zostaną także wyniki badań prowadzonych w zespole Ucznia Maszynowego Katedry Systemów i Sieci Komputerowych Politechniki Wroclawskiej.