

System wspierania oceny stanu ekologicznego ekosystemów jeziornych wspomagany algorytmami uczenia maszynowego

Streszczenie pracy

Wraz z ratyfikacją Ramowej Dyrektywy Wodnej w roku 2000 koncepcja ochrony ekosystemów jeziornych na stałe zadomowiła się wśród najważniejszych dla wspólnoty Europejskiej tematów środowiskowych. Natomiast ocena ich jakości dała możliwość formalnego sprawdzenia, czy przy jednoczesnym rozwoju społeczno-gospodarczym państw członkowskich zachowana jest zadowalająca kondycja ekologiczna wód śródlądowych. Niniejsza praca poświęcona jest fragmentowi procesu oceny wód jeziornych, a mianowicie ocenie ich stanu ekologicznego. Głównym celem badań było opracowanie systematycznego wsparcia oceny z wykorzystaniem algorytmów uczenia maszynowego. Automatyzacja i powtarzalność w konstruowaniu wskaźników stanu ekologicznego stanowi odpowiedź na potrzeby związane ze wsparciem procesów decyzyjnych w trakcie oceny, podczas priorytetyzacji działań ochronnych, a także prowadzeniem monitoringu osiągnięcia celów środowiskowych. System składa się z trzech współzależnych komponentów proponując rozwiązania w zakresie: imputacji danych brakujących, modelowania procesu oraz priorytetyzacji ekosystemów wskazanych do działań naprawczych i/lub ochronnych. Do kalibracji i testów wykorzystano dane pochodzące z drugiego cyklu planistycznego aktualizacji Planów Gospodarowania Wodami (2010-2015) pozyskane dla 499 jezior w Polsce. W pierwszym segmencie systemu zaproponowane zostało rozwiązanie problemów związanych z brakami w danych służących do konstruowania wskaźników oceny ekologicznej. Niedostatki w danych prowadzą do zaniechania oceny lub do wykorzystywania metod imputacji opartych jedynie na ich dostępności i popularności. Proponowana metoda opiera się na kombinacji algorytmów uczenia maszynowego z grupowaniem hierarchicznym. Dało to możliwość uzupełnienia braków w procesie imputacji semi-nadzorowanej. Położono również nacisk na interakcję eksperta ds. oceny ekologicznej z analitykiem danych, tak aby już na wczesnym etapie pracy z danymi uzyskać optymalny zbiór do stworzenia modelu oceny. W osi systemu znalazł się klasyfikator stanu ekologicznego oparty na maszynie wektorów nośnych korzystającej z jądra radialnej funkcji bazowej (kSVM). Do klasyfikacji wykorzystano wyniki pochodzące z zakończonej ewaluacji w połączeniu z danymi pomiarowymi dla dostępnych zmiennych służących do konstrukcji wskaźników stanu ekologicznego jezior. Klasyfikacji dokonano w przestrzeni zredukowanej do dwóch wymiarów wiodących z wykorzystaniem kombinacji dyskryminant liniowych. Pozwoliło to na odtworzenie oceny eksperckiej ze skutecznością przekraczającą 94%. Przygotowany model uodporniony został na występowanie zmian w rozkładach zmiennych wejściowych dzięki zastosowaniu optymalizacji hiperparametrów klasyfikatora kSVM. Uzyskany model wykorzystany został do stworzenia narzędzia wspierającego decyzje o priorytetyzacji ocenionych jezior w ramach klas stanu ekologicznego. W rozwiązaniu użyto nienadzorowanego algorytmu Kohonena do grupowania jezior tworząc scenariusze wyników priorytetyzacji, które następnie zmapowane zostały na rzeczywiste wyniki klasyfikacji dokonane przez ekspertów. Pozwoliło to na wskazanie miejsc, gdzie w ramach jednej z klas stanu ekologicznego tworzą się podgrupy, które nie były możliwe do identyfikacji w oryginalnym zestawie ocen. Jeziora w każdym z nowo powstałych klastrów zostały zaopatrzone w informację o swojej pozycji w zarówno w odniesieniu do oryginalnej klasyfikacji jak i w scenariuszach alternatywnego podziału. Dzięki temu zabiegowi priorytetyzacja mogła odbywać się zachowaniem struktury klas stanu ekologicznego wymaganej zapisami RDW. System został zaopatrzony w propozycje wizualizacji danych na każdym z etapów pracy z danymi. Komunikacja wizualna jest kluczowa nie tylko na linii ekspert - odbiorca. Również w przypadku interakcji w zespołach oceniających stan ekologiczny, odpowiedni sposób przedstawienia wyników kolejnych etapów może znacznie wspomóc zrozumienie nawet najbardziej złożonych treści. Toteż, głównymi adresatami niniejszej pracy są eksperci ds. oceny stanu ekologicznego śródlądowych wód powierzchniowych, specjaliści ds. wdrażania zapisów RDW, a także wspomagający ich analitycy danych.

Słowa kluczowe: systemy wsparcia decyzji, uczenie maszynowe, stan ekologiczny jezior