



**POLSKA AKADEMIA NAUK**  
**Instytut Badań Systemowych**

**BADANIA OPERACYJNE I SYSTEMOWE:  
ŚRODOWISKO NATURALNE,  
PRZESTRZEŃ, OPTYMALIZACJA**

**Olgierd Hryniewicz,  
Andrzej Straszak,  
Jan Studziński  
red.**



**BADANIA OPERACYJNE  
I SYSTEMOWE:  
ŚRODOWISKO NATURALNE, PRZE-  
STRZEŃ, OPTYMALIZACJA**

INSTYTUT BADAŃ SYSTEMOWYCH • POLSKA AKADEMIA NAUK

**Seria: BADANIA SYSTEMOWE**  
**tom 63**

---

**Redaktor naukowy:**

**Prof. dr hab. inż. Jakub Gutenbaum**

**Warszawa 2008**

**Olgierd Hryniewicz, Andrzej Straszak, Jan Studziński**

**BADANIA OPERACYJNE I SYSTEMOWE:  
ŚRODOWISKO NATURALNE, PRZESTRZEŃ,  
OPTIMALIZACJA**

Publikacja była opiniowana do druku przez zespół recenzentów, którego skład podano w treści tomu

Opinie, wyrażone przez autorów w pracach, zawartych w niniejszym tomie, nie są oficjalnymi opiniami Instytutu Badań Systemowych PAN, ani Polskiego Towarzystwa Badań Operacyjnych i Systemowych.

Copyright © by Instytut Badań Systemowych PAN & Polskie Towarzystwo Badań Operacyjnych i Systemowych  
Warszawa 2008

**ISBN 83-894-7519-7**  
**EAN 9788389475190**

Redakcja i opracowanie techniczne: Jan W. Owskiński, Aneta M. Pielak, Anna Gostyńska

**Lista recenzentów  
artykułów, wchodzących w skład tomów serii „Badania Systemowe”  
związanych z konferencją BOS 2008**

Dr Paweł Bartoszczuk  
Dr inż. Lucyna Bogdan  
Dr hab. inż. Zbigniew Buchalski  
Mgr inż. Hanna Bury  
Prof. dr hab. Marian Chudy  
Dr Jan Gadomski  
Mgr Grażyna Grabowska  
Mgr inż. Andrzej Jakubowski  
Dr hab. inż. Ignacy Kaliszewski  
Dr Andrzej Kałużko  
Dr hab. Leszek Klukowski  
Dr hab. inż. Wiesław Krajewski  
Dr inż. Lech Kruś  
Dr hab. inż. Marek Libura  
Dr Barbara Maźbic-Kulma  
Dr inż. Edward Michalewski  
Dr inż. Jan W. Owiński  
Dr inż. Grażyna Petriczek  
Dr inż. Henryk Potrzebowski  
Dr Maciej Romaniuk  
Prof. dr hab. Piotr Sienkiewicz  
Dr hab. Henryk Spustek  
Prof. dr hab. Andrzej Straszak  
Dr hab. inż. Jan Studziński  
Prof. dr hab. Tomasz Szapiro  
Mgr Anna Szediw  
Dr inż. Grażyna Szkatuła  
Dr hab. inż. Tadeusz Witkowski  
Dr Irena Woroniecka-Leciejewicz  
Dr hab. Sławomir Zadrozny  
Dr inż. Andrzej Ziółkowski

**Komitety Konferencji  
Badania Operacyjne i Systemowe 2008  
Rembertów, Akademia Obrony Narodowej**

Patronat honorowy

Bogdan Klich, Minister Obrony Narodowej  
Maciej Nowicki, Minister Środowiska i Zasobów Naturalnych

Komitet Sterujący

Janusz Kacprzyk, Prezes Polskiego Towarzystwa Badań Operacyjnych i Systemowych  
Olgierd Hryniewicz, Dyrektor Instytutu Badań Systemowych  
Janusz Kręcikij, Komendant Akademii Obrony Narodowej

Komitet Programowy

Piotr Sienkiewicz, *Przewodniczący*  
Jacek Mercik, *Wiceprzewodniczący*

<i>Tomasz Ambroziak</i>	<i>Ryszard Budziński</i>	<i>Wojciech Cellary</i>
<i>Marian Chudy</i>	<i>Ludostaw Drelichowski</i>	<i>Jerzy Hołubiec</i>
<i>Olgierd Hryniewicz</i>	<i>Adam A. Janiak</i>	<i>Jerzy Józefczyk</i>
<i>Ignacy Kaliszewski</i>	<i>Józef Korbicz</i>	<i>Maciej Krawczak</i>
<i>Piotr Kulczycki</i>	<i>Małgorzata Łatuszyńska</i>	<i>Marek J. Malarski</i>
<i>Barbara Mażbic-Kulma</i>	<i>Zbigniew Nahorski</i>	<i>Andrzej Najgebauer</i>
<i>Włodzimierz Ogryczak</i>	<i>Wojciech Olejniczak</i>	<i>Jan W. Owsiański</i>
<i>Andrzej Piegat</i>	<i>Krzysztof Santarek</i>	<i>Roman Słowiński</i>
<i>Honorata Sosnowska</i>	<i>Henryk Spustek</i>	<i>Jan Stachowicz</i>
<i>Andrzej Straszak</i>	<i>Tomasz Szapiro</i>	<i>Andrzej Szymonik</i>
<i>Ryszard Tadeusiewicz</i>	<i>Eugeniusz Toczyłowski</i>	<i>Tadeusz Trzaskalik</i>
<i>Jan Węglarz</i>	<i>Tadeusz Witkowski</i>	<i>Stanisław Zajas</i>
	<i>Bogdan Zdrowski</i>	

Komitet Organizacyjny

Jan W. Owsiański, Andrzej Kałużko, Mieczysław Pelc, Zbigniew Piątek

Sekretariat

Krystyna Warzywoda, Monika Majkut, Aneta M. Pielak, Krzysztof Sep,  
Anna Stachowiak, Halina Świeboda, Tadeusz Winiarski

Redakcja wydawnictw

Janusz Kacprzyk, Piotr Sienkiewicz, Andrzej Najgebauer,  
Olgierd Hryniewicz, Andrzej Straszak, Jan Studziński,  
Jan W. Owsiański, Zbigniew Nahorski, Tomasz Szapiro

# **Środowisko i jego ochrona**



## **PORÓWNANIE KLASYFIKATORÓW DLA ISW ZARZĄDZANIA SIECIĄ WODOCIĄGOWĄ<sup>11</sup>**

**Izabela Rojek**

Uniwersytet Kazimierza Wielkiego, Instytut Mechaniki Środowiska  
i Informatyki Stosowanej, ul. Chodkiewicza 30, 85-064 Bydgoszcz

W artykule przedstawiono porównanie modeli klasyfikacyjnych dla inteligentnych systemów wspomagania (ISW) zarządzania siecią wodociągową. Modele pojedynczych klasyfikatorów i wielokrotnych zostały porównane dla uzyskania optymalnej klasyfikacji. Modele zostały przetestowane na danych rzeczywistych z miejskiego przedsiębiorstwa wodno-kanalizacyjnego. Bazując na modelach klasyfikacji inteligentny system wspomagania pozwala na tworzenie scenariuszy dla zarządzania siecią wodociągową. Dlatego wbudowane modele ulepszają zarządzanie siecią wodociągową. Przyszłe badania skoncentrowane zostaną na tworzeniu klasyfikatorów hybrydowych integrujących różne pojedyncze modele już w fazie uczenia.

### 1. Wstęp

Idea zintegrowanego systemu informatycznego wspomagania decyzji dla miejskiego systemu wodno-ściekowego powstała w wyniku obserwacji, że poszczególne obiekty zarządzane przez typowe przedsiębiorstwo wodociągowe nie funkcjonują niezależnie, a są elementami jednego układu wzajemnie na siebie wpływającymi. Funkcję centralną pełni w tym układzie sieć wodociągowa a jej zmieniające się w czasie obciążenie wpływa na działanie zestawów pompowych w stacji ujęcia wody, na obciążenie hydrauliczne sieci kanalizacyjnej i w rezultacie na jakość pracy oczyszczalni ścieków. Poprawne przewidywanie obciążenia i sterowanie operacyjne siecią wodociągową pozwoli energooszczędnie sterować pompami na ujęciu wody oraz efektywnie sterować procesem technologicznym w oczyszczalni, przygotowując ją odpowiednio wcześniej do przyjęcia określonej ilości ścieków i ładunku zanieczyszczeń (Studziński i Bogdan, 2006).

W skład zintegrowanego miejskiego systemu wodno-ściekowego wchodzi: stacja ujęcia i uzdatniania wody, sieć wodociągowa, sieć kanalizacyjna i oczyszczalnia ścieków. Założeniem jest, że te cztery moduły będą z sobą współpracowały oraz mogą również być wykorzystywane jako autonomicznie działające systemy. Częścią integrującą te systemy jest branżowa baza danych, zawierająca dane techniczne, technologiczne i eksploatacyjne o wszystkich obiektach systemu wodno-ściekowego. W systemie będą wykorzystane modele matematyczne badanych obiektów, które posłużą do obliczeń symulacyjnych i prognozowania oraz algorytmy optymalizacji, wykorzystane do usprawniania funkcjonowania obiektów i generowania scenariuszy postępowania w stanach normalnej eksploatacji i w sytuacjach awaryjnych.

---

<sup>11</sup> Praca wykonana w ramach projektu badawczego MNiSzW nr R11 001 01.

Pierwszy etap tworzenia całego systemu dotyczy inteligentnego systemu wspomagania zarządzania siecią wodociągową. System inteligentnego zarządzania sprawdza czy bieżące mierzone parametry ciśnienia i przepływu wody mieszczą się w granicy normy. Na początku prowadzone były badania wykorzystujące prosty klasyfikator – metodę indukcji drzew decyzyjnych (Rojek, 2007).

W artykule przedstawiono badania dotyczące opracowania pojedynczych klasyfikatorów oraz złożonych klasyfikatorów oraz porównania tych modeli dla uzyskania optymalnej klasyfikacji. Modele zostały przetestowane na danych rzeczywistych z miejskiego przedsiębiorstwa wodno-kanalizacyjnego. Opracowane klasyfikatory pozwolą na ciągłe kontrolowanie, czy przepływ i ciśnienie wody w sieci wodociągowej są w normie. W przypadku nieprawidłowości zostaną uruchomione odpowiednie procedury dotyczące pracy pomp.

Zintegrowany system informatyczny wspomagania decyzji dla miejskiego przedsiębiorstwa wodno-ściekowego powstaje w ramach projektu badawczego finansowego przez Ministerstwo Nauki i Szkolnictwa Wyższego nr R11 001 01.

## 2. Przykład pojedynczego klasyfikatora – drzewo decyzyjne

Metoda indukcji drzew decyzyjnych pozwala na przybliżenie funkcji klasyfikacyjnych o dyskretnych wartościach wejściowych odnoszących się do pewnych pojęć, klas decyzyjnych.

Podstawowym algorytmem klasyfikacji są drzewa decyzyjne (ang. decision trees), w których wewnętrzne węzły oznaczają test atrybutu, gałęzie reprezentują wyniki testów, a liście reprezentują etykiety klas. Drzewa decyzyjne są budowane metodą „z góry na dół”. Umożliwiają one rozwiązywanie problemów, w których wymagana jest klasyfikacja przypadków na wiele kategorii. Narzędzia wspomagające analizę za pomocą drzew decyzyjnych generują statystyki klasyfikacji, diagram drzewa decyzyjnego oraz tzw. tablicę CM (ang. *confusion matrix*), czyli tablicę reprezentującą poprawnie i niepoprawnie zaklasyfikowane przypadki (Hand i in., 2005; Larose, 2006). Do oceny drzewa decyzyjnego jako systemu klasyfikującego użyto kryterium oceny błędu klasyfikowania, które jest podstawowym narzędziem weryfikacji. Jest on zdefiniowany jako (1):

$$\mathcal{E}_{ov} = \frac{n_w}{n_{test}}. \quad (1)$$

$n_w$  – liczba błędnie sklasyfikowanych przykładów testowych,  $n_{test}$  – liczba przykładów testowych.

Weryfikację drzewa przeprowadza się w odniesieniu do zbioru przykładów testowych, które nie były częścią zbioru uczącego. Przykłady testowe są klasyfikowane za pomocą klasyfikatora powstałego ze zbioru uczącego. Ponadto, dla przykładu testowego znana jest jego rzeczywista klasyfikacja. Pozwala to na porównanie

proponowanej przez klasyfikator decyzji z rzeczywistością i stwierdzenie, czy decyzja klasyfikatora jest poprawna, czy błędna.

Na podstawie klasyfikatora w postaci drzewa zostały automatycznie wygenerowane reguły decyzyjne, które zostały umieszczone w bazie wiedzy systemu ekspertowego.

### 3. Złożone klasyfikatory – klasyfikatory wielokrotne

W ostatnich latach obserwuje się rosnące zainteresowanie tworzeniem złożonych systemów klasyfikujących. Idea ta polega na integracji wielu pojedynczych algorytmów uczenia się w jeden system klasyfikujący. Celem takiej integracji jest przede wszystkim osiągnięcie lepszej trafności klasyfikacji od otrzymanej w rezultacie użycia oddzielnie pojedynczych klasyfikatorów wchodzących w skład systemu (Stefanowski, 2001). W latach dziewięćdziesiątych powstało wiele prac dotyczących złożonych systemów klasyfikujących. Wyniki eksperymentalne potwierdzają wzrost trafności klasyfikowania dla proponowanych systemów. Integracja jednorodnych lub różnorodnych klasyfikatorów (lub algorytmów uczenia się) może być przeprowadzona różnymi sposobami.

Jedną z dwóch głównych kategorii złożonych systemów klasyfikujących są systemy wielokrotne. Pojęcie systemu wielokrotnego (odpowiednik terminu angielskiego *multiple model*, *multiple classifier* lub *ensemble of classifiers*) odnosi się do zbioru pojedynczych klasyfikatorów, których odpowiedzi są zagregowane do jednej odpowiedzi całego systemu. Klasyfikatory składowe mogą być jednorodne (homogeniczne) lub różnorodne (heterogeniczne). Zasadnicze pytanie dotyczy sposobów budowania złożonych systemów, tak aby pozwalały na osiągnięcie wyższej trafności klasyfikacji niż klasyfikatory składowe działające oddzielnie.

Zgodnie z opinią niektórych badaczy klasyfikatory powinny działać z pewnym stopniem niezgodności, tzn. popełniane przez nie błędy klasyfikowania powinny być wzajemnie nieskorelowane. Rozważania teoretyczne nad wielokrotnymi systemami klasyfikującymi dla problemów dwuklasowych wskazują, że jeśli błędy klasyfikatorów składowych nie są ze sobą skorelowane i każdy klasyfikator popełnia błąd poniżej 50%, to całkowity błąd systemu maleje proporcjonalnie do liczby użytych klasyfikatorów (Stefanowski, 2001).

Agregacja odpowiedzi pojedynczych klasyfikatorów w globalną odpowiedź systemu dokonywana jest w sposób grupowy albo specjalizowany. W pierwszym sposobie wszystkie klasyfikatory biorą udział w wypracowaniu decyzji końcowej dla nowego obiektu. W drugim sposobie wyznacza się te klasyfikatory, których specjalizacja obejmuje charakterystykę klasyfikowanego obiektu. Sama agregacja realizowana jest najczęściej poprzez głosowanie proste (głos każdego klasyfikatora ma równy priorytet) lub ważne. Istnieje wiele sposobów tworzenia złożonych systemów klasyfikujących. Część z nich polega na modyfikowaniu danych uczących dla poszczególnych jednorodnych klasyfikatorów, inne zakładają różnicowanie samych modeli klasyfikujących.

Niewiele jest dotychczas prac przeglądowych dotyczących różnych propozycji. Poniżej zamieszczono skrótkowo klasyfikację podaną w Stefanowski (2001):

Systemy wielokrotne dzielą się na klasyfikatory oparte na:

- zróżnicowanych klasyfikatorach
  - o Generalizacja stosowa (Stacking) - model hierarchicznego, wielowarstwowego klasyfikatora złożonego zaproponowany przez Wolperta. Klasyfikatory składowe pierwszej warstwy uczone są na oryginalnych danych, a ich odpowiedzi są danymi wejściowymi do kolejnej warstwy. W ostatniej warstwie umieszczony jest pojedynczy klasyfikator wypracowujący ostateczną decyzję. Klasyfikatory wyższych warstw starają się minimalizować błędy popełniane przez klasyfikatory z niższych warstw.
  - o Technika bagging - model wprowadzony przez Breimana oparty na idei powielania zbioru uczącego za pomocą wielokrotnego próbkowania techniką "bootstrapping". Każdy powielony zbiór ma tę samą wielkość, co oryginalny, lecz w wyniku próbkowania niektóre przykłady uczące są kopiowane, podczas gdy inne nie mogą się wielokrotnie powtórzyć. Z każdego z tych zbiorów generuje się klasyfikator tym samym algorytmem uczącym. Odpowiedzi klasyfikatorów są agregowane poprzez proste głosowanie.
  - o Technika "boosting" - podana przez Freund'a i Schapirego; polega na skojarzeniu z każdym obiektem uczącym wagi, której wartość odzwierciedla jego istotność. Algorytm uczenia się uruchamiany jest iteracyjnie nad zmodyfikowanym zbiorem uczącym. Modyfikacje dotyczą zmiany wag w kolejnych kopiach zbioru uczącego, które skupiają działanie algorytmu uczącego wokół tzw. trudnych przykładów, dla których klasyfikatory popełniały błędy w poprzednich iteracjach. Odpowiedź końcowa systemu opiera się na ważonym głosowaniu (tutaj wagi głosujących klasyfikatorów uzależnione są od błędów popełnianych przez odpowiednie klasyfikatory składowe).

Porównanie eksperymentalne złożonych klasyfikatorów zbudowanych za pomocą technik "bagging" oraz "boosting" przedstawiono w Quinlan (1996).

Według autora techniki te są skuteczne, jeśli używa się tzw. niestabilnych algorytmów uczących się, tzn. takich, które tworzą różne klasyfikatory w odpowiedzi na małe zmiany w zbiorze uczącym.

#### 4. Problem zarządzania siecią wodociągowej

Przedsiębiorstwo wodociągowe zajmuje się, m.in., dystrybucją wody dobrej jakości w ilości gwarantującej zaspokojenie potrzeb odbiorców, poprawną eksploatacją sieci wodociągowej zapewniającą właściwe ciśnienie w węzłach odbiorczych, sprawnym usuwaniem awarii oraz planowaniem i wykonywaniem prac związanych z konserwacją, modernizacją i rozbudową sieci (Studziński i Bogdan, 2006). Zarządzanie siecią wodociągową jest trudnym i kompleksowym procesem, ze względu na dynamikę i stochastyczność zmian zachodzących w sieci. Tego typu zmiany wymuszają zastosowanie metod sztucznej inteligencji.

Problemy zarządzania siecią wodociągową dotyczą optymalizacji jej struktury i parametrów, energooszczędnej eksploatacji zapewniającej poprawną dystrybucję wody o odpowiedniej jakości, optymalizacji prac projektowych przy rozbudowie sieci, wykrywania i lokalizacji stanów awaryjnych, opracowywania planów remontów sieci z uwzględnieniem jej awaryjności, kosztów inwestycyjnych i eksploatacyjnych. W szczególności, ważnym problemem jest energooszczędne zarządzanie pompami w sieci oraz kontrolowanie poziomu wody w zbiornikach retencyjnych. Na początku prowadzone były badania dotyczące analizy, czy ciśnienie i przepływ wody w sieci wodociągowej mieszczą się w normie, czy też nie. W tym celu opracowano system ekspertowy, który sygnalizuje potrzebę uruchomienia pomp.

## 5. Modele klasyfikacji dla ISW zarządzania siecią wodociągową

### 5.1. Przygotowanie danych

Badania zostały przeprowadzone dla wybranych węzłów sieci wodociągowej: osiedle domków jednorodzinnych, węzły dla 4 hydroforni z różnych stref ciśnieniowych sieci wodociągowej. Mierzone dane przedstawiają stan sieci wodociągowej w tych wybranych mierzonych węzłach sieci. Pomiary robione były co 10 s. Dane zostały przesłane do bazy danych przy użyciu sieci PlusGSM. Następnie z bazy danych informacje dotyczące ciśnienia i przepływu wody posłużyły do utworzenia pliku uczącego. Plik ten składa się z:

- atrybutów wejściowych: dzień tygodnia, dzień miesiąca, miesiąc, czas, przepływ i ciśnienie,
- atrybutu wyjściowego: decyzji, czy wartość ciśnienia i przepływu jest w normie, czy też nie.

Plik zawiera dane z jednego miesiąca, co stanowi 65534 przypadków. Jest przygotowany w formacie ARFF.

### 5.2. Opis modeli klasyfikacji

W eksperymencie przeprowadzono badania dla klasyfikatorów prostych i wielokrotnych. Użyto następujących pojedynczych klasyfikatorów: *Decision Stump*, *J48*, *Random Tree*, *REPTree*, *ZeroR* i klasyfikatorów wielokrotnych: *AdaBoost*, *LogitBoost*, *RacedIncrementalLogitBoost*, *FilteredClassifier*, *Bagging*, *Stacking*, *AttributeSelectedClassifier*, *OrdinalClassClassifier*, *RandomCommittee*.

Do oszacowania skuteczności klasyfikatorów wykorzystano technikę 10-krotnej oceny krzyżowej. W tej technice klasyfikowania trafności zbiorów przykładów jest losowo dzielony na  $k$  równolicznych podzbiorów  $U = E_1 \cup \dots \cup E_k$ . W  $i$ -tej iteracji ( $1 \leq i \leq k$ ) zbiór  $E^U = U/E_i$  jest stosowany jako zbiór uczący, a sam zbiór  $E_i$  jako zbiór przykładów testowych. Trafność klasyfikowania jest wyliczana jako średnia z trafności estymowanych w każdej iteracji  $\eta(E_i)$ , tzn. jest zdefiniowana jako:

$$\eta_{ov} = \frac{1}{k} \sum_{i=1}^k \eta(E_i). \quad (2)$$

W ogólności, dobór parametru  $k$  powinien być dokonany w zależności od rozmiarów analizowanego zbioru danych. Najbardziej zalecaną wartością jest  $k=10$ , choć dla mniejszych zbiorów danych może być ona większa (Stefanowski, 2001).

Klasyfikatory *AdaBoostM1*, *LogitBoost* i *RacedIncrementalLogitBoost* używają *weka.classifiers.trees.DecisionStump*. Klasyfikator **OrdinalClassClassifier** używa *weka.classifiers.trees.J48*. Klasyfikator *AttributeSelectedClassifier* używa *weka.attributeSelection.CfsSubsetEval*, *weka.attributeSelection.BestFirst* oraz *weka.classifiers.trees.J48*. Klasyfikator *FilteredClassifier* używa *weka.classifiers.trees.J48* na danych filtrowanych przez *weka.filters.supervised.attribute.Discretize -R first-last*. Klasyfikator *Bagging* używa *weka.classifiers.trees.REPTree*. Klasyfikator *Stacking* używa *weka.classifiers.rules.ZeroR*. Klasyfikator **RandomCommittee** używa *weka.classifiers.trees.RandomTree*.

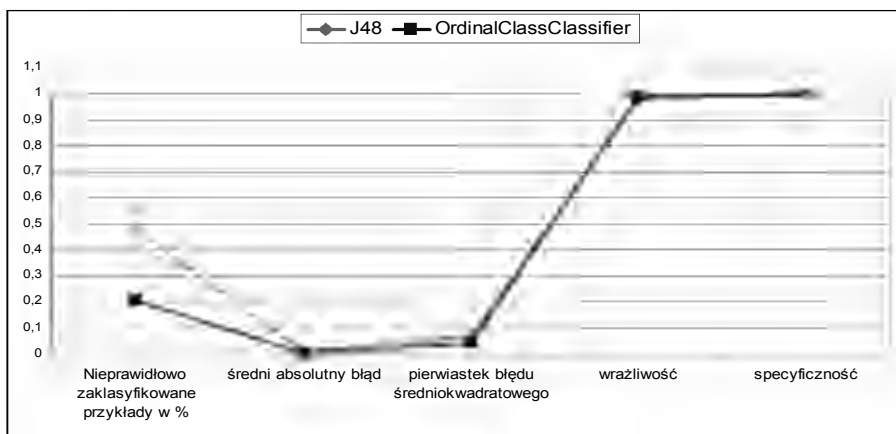
Rys. 1 prezentują tabelę, w której pokazano porównanie wybranych klasyfikatorów ze względu na:

- prawidłowo zaklasyfikowane przykłady w %,
- średni błąd absolutny,
- pierwiastek błędu średniokwadratowego,
- wrażliwość,
- specyficzność.

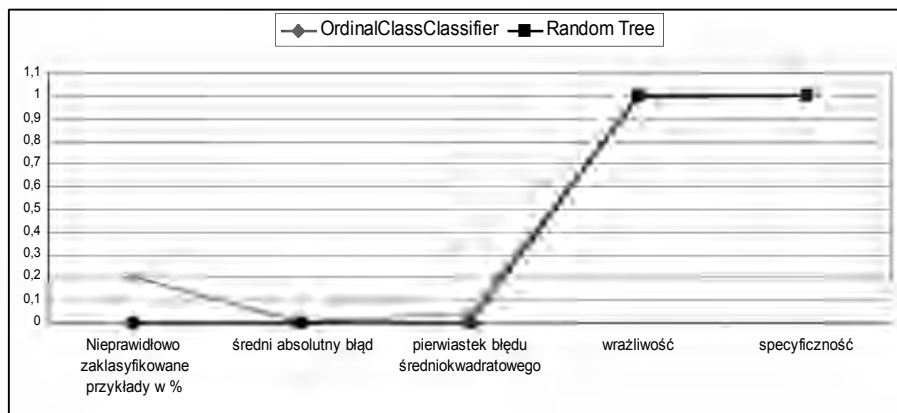
Rys. 1. Analiza i porównanie klasyfikatorów.

Nazwa klasyfikatora	Prawidłowo zaklasyfikowane przykłady w %	Średni błąd absolutny	Pierwiastek błędu średniokwadratowego	Wrażliwość	Specyficzność
Decision Stump	85,8547	0,1843	0,3036	0	1
AdaBoost	89,9182	0,1637	0,285	0,393	0,983
LogitBoost	90,0006	0,1507	0,2721	0,405	0,982
Raced Incremental Logit Boost	90,3363	0,1345	0,2633	0,485	0,972
J48	99,5178	0,0066	0,0675	0,978	0,998
Attribute Selected Classifier	95,6526	0,0697	0,1867	0,781	0,986
Filtered Classifier	95,7274	0,0648	0,1801	0,803	0,983
<b>Ordinal Class Classifier</b>	<b>99,7955</b>	<b>0,0038</b>	<b>0,0438</b>	<b>0,989</b>	<b>0,999</b>
REPTree	99,3332	0,0098	0,0783	0,971	0,997
Bagging	99,5025	0,0102	0,0633	0,976	0,998
ZeroR	85,8547	0,2429	0,3485	0	1
Stacking	85,8547	0,2429	0,3485	0	1
<b>Random Tree</b>	<b>100</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>1</b>
<b>Random Committee</b>	<b>100</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>1</b>

Następnie przedstawiono wykresy, które porównują wybrane klasyfikatory pojedyncze z wielokrotnymi. Rys. 2 przedstawia porównanie klasyfikatora pojedynczego *J48*, który jest implementacją ostatniej publicznie dostępnej wersji algorytmu indukcji drzew decyzyjnych *C4.5* z klasyfikatorem wielokrotnym *OrdinalClassClassifier*, który używa *J48*. Rys. 3 pokazuje porównanie klasyfikatora wielokrotnego *OrdinalClassClassifier* z pojedynczym klasyfikatorem *Random Tree*, który okazał się najlepszy.



Rys. 2. Porównanie klasyfikatora *J48* z *OrdinalClassClassifier*



Rys. 3. Porównanie klasyfikatora *OrdinalClassClassifier* z *Random Tree*

## 6. Podsumowanie

Ogólna zasada stosowana w nauce głosi, że w przypadku gdy istnieje możliwość wyboru pomiędzy modelem prostym i bardziej złożonym, należy zawsze preferować model prostszy - o ile oczywiście ten drugi nie dopasowuje się znacząco lepiej do posiadanych danych.

Według tej zasady należy również postępować w przypadku klasyfikatorów. Analizując klasyfikatory najlepszymi okazały się klasyfikatory wykorzystujące algorytm Random Tree. Nieco gorszy okazał się klasyfikator wielokrotny OrdinalClassClassifier wykorzystujący J48. Te klasyfikatory najlepiej nadają się do zastosowania w inteligentnym systemie wspomagania zarządzania siecią wodociągową.

W przypadku porównywalnego dopasowywania się tych klasyfikatorów przy wyborze typu klasyfikatora do tworzenia modelu klasyfikacyjnego, należy również kierować się złożonością samego modelu, rozpatrując czas tworzenia klasyfikatora, obliczeń i dokładności.

Przyszłe badania skoncentrowane zostaną na tworzeniu klasyfikatorów hybrydowych integrujących różne pojedyncze modele już w fazie uczenia.

## Literatura

- Hand D., Mannila H., Smith P. (2005) *Eksploracja danych*. WNT, Warszawa.
- Larose D. T. (2006) *Odkrywanie wiedzy z danych*. Wprowadzenie do eksploracji danych. PWN, Warszawa.
- Quinlan J.R. (1996) Bagging, boosting and C4.5. W: *Proceedings of the 13<sup>th</sup> National Conference on Artificial Intelligence*, 725-730.
- Rojek I. (2007) Inteligentny system wspomagania decyzji dla sterowania siecią wodociągową. W: *Materiały konferencyjne II Krajowej Konferencji Naukowej „Technologie przetwarzania danych”*, wydawnictwo Politechniki Poznańskiej, Poznań, 186-194.
- Stefanowski J. (2001) *Algorytmy indukcji reguł decyzyjnych w odkrywaniu wiedzy*. Wydawnictwo Politechniki Poznańskiej, seria Rozprawy, **361**, Poznań.
- Studziński J., Bogdan L. (2006) Informatyczny system wspomagania decyzji do zarządzania, sterowania operacyjnego i planowania miejskiego systemu wodno-ściekowego. W: J. Studziński, L. Drelichowski, O. Hryniewicz, red., *Rozwój i zastosowania metod ilościowych i technik informatycznych wspomagających procesy decyzyjne*, Instytut Badań Systemowych PAN, *Badania Systemowe*, **49**, Warszawa, 149-157.



IBS PAN *Konf.*

46003

Bibl. podręczna

**Olgierd Hryniewicz, Andrzej Straszak, Jan Studziński  
red.**

**BADANIA OPERACYJNE I SYSTEMOWE:  
ŚRODOWISKO NATURALNE, PRZESTRZEŃ,  
OPTYMALIZACJA**

Książka składa się z artykułów przedstawiających wyniki prac z dziedziny badań operacyjnych i systemowych, poświęconych środowisku naturalnemu i zarządzaniu nim, zwłaszcza w zakresie ochrony atmosfery, globalnego ocieplenia i walki z nim, jakości i zaopatrzenia w wodę. Tematyka ta jest rozszerzona o aspekty przestrzenne, regionalne i samorządowe, a także planowanie i funkcjonowanie infrastruktury. Tom zamykają prace metodyczne, dostarczające technik, będących podstawą prezentowanych zastosowań.

**ISBN 83-894-7519-7**

**EAN 9788389475190**

---

---

**Instytut Badań Systemowych PAN**

**tel. (4822) 3810241 / 3810273 e-mail: [biblioteka@ibspan.waw.pl](mailto:biblioteka@ibspan.waw.pl)**