



System ekspertowy do oceny układu systemu dystrybucji wody sporządzony za pomocą wnioskowania indukcyjnego

Jacek Dawidowicz
Politechnika Białostocka

1. Wstęp

Programy do symulacji komputerowych stosowane są powszechnie od wielu lat w modelowaniu systemów dystrybucji wody [5, 15]. Obecnie programy symulacyjne wyposażone są w wiele usprawnień w zakresie możliwości wprowadzania danych, analizy wyników, udogodnień interfejsu graficznego [6, 19]. Programy do symulacji systemów dystrybucji wody coraz częściej korzystają z możliwości GIS [21, 22] oraz CAD [23]. Oczekiwania w stosunku do programów komputerowych przeznaczonych do symulacji hydraulicznych nie ograniczają się jednak obecnie wyłącznie do podstawowych zadań. Coraz częściej dąży się do stworzenia programów komputerowych, które będą charakteryzowały się, chociaż w niewielkim stopniu pracą kreatywną, wspomagały projektanta lub operatora w podejmowaniu decyzji. Narzędzia informatyczne, które mogą być zastosowane do realizacji powyższych zadań należą metod inteligencji obliczeniowej [8, 20]. W zakres powyższego podejścia wchodzi m.in. systemy ekspertowe.

2. Systemy ekspertowe

Systemy ekspertowe (ang. *expert systems*) są programami komputerowymi, które na podstawie zgromadzonej wiedzy mogą wyciągać wnioski i podejmować decyzje, działając w sposób zbliżony do rozumowania człowieka [2, 3, 10, 14].

Systemy ekspertowe zbudowane są z oddzielnych, współpracujących ze sobą modułów. W jednym z nich znajduje się sformalizowana,

odseparowana od pozostałej części systemu wiedza, dotycząca określonej dziedziny. Ta część systemu nazywana jest *bazą wiedzy* (ang. *knowledge base*). Inny moduł zawiera procedury wnioskowania, czyli tzw. *maszynę wnioskującą* lub inaczej *kontroler wyводу* (ang. *inference engine*). Ważnym elementem są procedury sterowania dialogiem, czyli *interfejs* umożliwiający komunikację użytkownika z programem. Systemy ekspertowe nie posiadają jawnego opisu sposobu rozwiązania określonego problemu, czyli algorytmu. Moduł wnioskujący rozwiązuje problem we współpracy z użytkownikiem, wykorzystując wiedzę z danej dziedziny, zapisaną w bazie wiedzy.

2.1. Reprezentacja wiedzy w systemach ekspertowych

Wiedza w potocznym znaczeniu jest zbiorem wiadomości z określonej dziedziny. Wiedza może być też rozumiana jako symboliczny opis otaczającego nas świata rzeczywistego, zawierający aksjomatyczne i empiryczne relacje występujące w tym świecie. W taki sposób pojęcie wiedzy rozumiane jest w problematyce systemów ekspertowych. W tym wypadku wyróżnia się dwa podstawowe typy symbolicznej reprezentacji wiedzy:

1. reprezentacja proceduralna, polegająca na zdefiniowaniu zbioru procedur, których działanie reprezentuje wiedzę z wybranej dziedziny,
2. reprezentacja deklaratywna, polegająca na określeniu zbioru specyficznych dla rozpatrywanej dziedziny faktów, stwierdzeń, reguł.

Reprezentacja wiedzy zapewnia zachowanie określonego formalizmu zapisu, przez co wiedza jest „zrozumiała” dla systemu ekspertowego. Istnieje wiele metod reprezentacji wiedzy deklaratywnej, w tym reguły, sieci semantyczne, formuły logiki predykatów oraz ramy. W większości zbudowanych dotychczas systemów ekspertowych wykorzystano regułową reprezentację wiedzy [1].

2.2. Pozyskiwanie wiedzy do systemów ekspertowych

Zasadniczym problemem przy budowie systemów ekspertowych jest proces *pozyskiwania wiedzy* (ang. *knowledge acquisition*). Od jakości utworzonej bazy wiedzy w znacznej mierze zależy poprawność i efektywność funkcjonowania systemu ekspertowego. Tradycyjnymi źródłami pozyskiwania wiedzy są konsultacje ze specjalistami z określonej dzie-

dziny, obserwacja i analiza pracy eksperta, ankiety oraz literatura fachowa. Akwizycja wiedzy przy użyciu powyższych metod jest jednak zadaniem trudnym i czasochłonnym. Coraz częściej stosuje się automatyczne metody pozyskiwania wiedzy, co ma ścisły związek z zagadnieniami uczenia maszynowego (systemów uczących się) (ang. *machine learning*) [4]. Pozyskiwanie wiedzy odbywa się w tym wypadku na zasadzie analizy zbioru przykładów reprezentujących określone zagadnienie.

2.3. Wnioskowanie indukcyjne

Najbardziej popularną strategią automatycznego nabywania wiedzy w uczeniu maszynowym jest wnioskowanie indukcyjne. Rozumiane jest jako przechodzenie od jednostkowych obserwacji, reprezentowanych przez przykłady, do wniosków ogólnych. Opierając się na przykładach próbuje się wykryć ogólne prawidłowości w rozpatrywanym zagadnieniu. Zbiór przykładów powinien być reprezentatywny i w zadowalający sposób opisywać charakterystyczne cechy danego zagadnienia. Przykłady mogą być otrzymywane różnymi metodami, m.in. poprzez opisanie poszczególnych faktów, obserwacji lub problemów przez eksperta, generowanie przez specjalistyczne programy symulacyjne lub w wyniku odczytów z eksploatacji określonego urządzenia, itp. [10]. Uczenie się indukcyjne określane jest również jako wnioskowanie „od szczegółu do ogółu”. Uogólnienie wiedzy na podstawie wybranych przykładów, na wszystkie możliwe obiekty danego zagadnienia określa się również mianem generalizacji. Pozwala to zastosować pozyskaną wiedzę do podejmowania decyzji przy nowych obserwacjach.

Wnioskowanie indukcyjne występuje w trzech odmianach, które w uczeniu maszynowym określane są jako: uczenie się pojęć, tworzenie pojęć i uczenie się aproksymacji funkcji. W niniejszej pracy wykorzystano metodę uczenia się pojęć. Termin *pojęcie* (ang. *concept*) rozumiany jest w uczeniu maszynowym jako „ogólne określenie oznaczające zbiór (klasę) obiektów posiadających pewne wspólne właściwości, które odróżniają je od innych pojęć” [9]. Pojęcia określają więc grupy obiektów, które nazywane są *kategoriami* lub *klasami*. Uczenie się pojęć można również nazwać problemem klasyfikacji, lecz realizowanym przez techniki uczenia maszynowego.

3.2. Indukcja drzew decyzyjnych

Najczęściej stosowanym algorytmem wnioskowania indukcyjnego jest generowanie drzew decyzyjnych. Metoda ta została opracowana przez R. Quinlana i nazywana w kolejnych wersjach ID3, C4, C4.5 [16–18]. Drzewem decyzyjnym nazywamy graf skierowany acykliczny. Początek grafu nazywany jest korzeniem, natomiast na końcu znajdują się tzw. liście, które odpowiadają poszukiwanym klasom. Pomędzy nimi znajdują się węzły, do których przypisane są testy pozwalające na przejście grafu od korzenia do jednego z liści, co wiąże się z wyborem odpowiedniej klasy. Testy na atrybutach w węzłach są rozłączne. Oznacza to, że dla każdego obiektu istnieje jedna, jednoznacznie określona ścieżka klasyfikacji. Obiekty są oceniane na podstawie wartości opisujących je atrybutów, zgodnie z kolejnością wynikającą ze struktury drzewa. Wybór ścieżki w drzewie decyzyjnym oparty jest na testach skonstruowanych z wykorzystaniem teorii entropii. W niniejszej pracy do indukcji drzewa decyzyjnego zastosowano system DeTreex, służący do wspomaganie procesu pozyskiwania wiedzy [13]. Dzięki zastosowanej indukcyjnej metodzie uczenia maszynowego możliwe jest budowanie drzew decyzyjnych na podstawie zbioru przykładów uczących. System DeTreex jest niezależny dziedzinowo i może być zastosowany w różnych dziedzinach nauki. System należy do pakietu sztucznej inteligencji Sphinx [11, 12].

4. Sporządzenie zbioru danych uczących do indukcji drzewa decyzyjnego

Zastosowanie metody indukcji drzew decyzyjnych pociąga za sobą konieczność przygotowania zbioru przykładów uczących. Bardzo ważne jest, aby zbiór uczący w pełni reprezentował populację, czyli dziedzinę problemu. W przypadku uczenia się pojęć (klasyfikacji) w zbiorze uczącym muszą znajdować się przykłady odpowiadające wszystkim klasom. Ponadto przykłady odnoszące się do poszczególnych klas powinny opisywać właściwą zmienność atrybutów wewnątrz każdej klasy. Wykorzystując informacje o 8 systemach wodociągowych i parametrach o ich pracy, przeprowadzono obliczenia hydrauliczne sieci wodociągowych. Ponadto dokonano korekty danych w celu uzyskania wariantów obliczeniowych sieci wodociągowych, wymagających zastosowania pompowni sieciowej lub reduktora ciśnienia. Część obliczeń hy-

draulicznych wykonano celowo dla niepoprawnych układów systemów dystrybucji wody, tak by system uczący się dysponował reprezentacją powyższych warunków w zbiorze uczącym.

Przykłady uczące odnoszą się do poszczególnych węzłów sieci wodociągowych i zostały opracowane w taki sposób, by mogły służyć do oceny układu dystrybucji wody. W celu przeprowadzenia indukcji drzewa decyzyjnego zdefiniowano atrybuty charakteryzujące układy systemów dystrybucji wody:

- wysokość ciśnienia w węźle zasilającym sieć (H_{zrodla}),
- długość najkrótszej drogi w sieci od węzła zasilającego do innego wybranego węzła (L_{min}),
- różnica wysokości terenu pomiędzy węzłem zasilającym a bieżąco wybranym węzłem sieci ($Roznica_wysokosci_terenu$),
- maksymalna różnica wysokości terenu występująca na najkrótszej drodze pomiędzy węzłem zasilającym a aktualnie wybranym węzłem sieci ($Maks_roznica_wys_terenu$),
- suma strat ciśnienia na najkrótszej drodze pomiędzy węzłem zasilającym a bieżąco wybranym węzłem sieci ($Suma_strat_cis$),
- maksymalna wysokość ciśnienia na najkrótszej drodze pomiędzy węzłem zasilającym a wybranym węzłem sieci ($Maks_wys_cis$),
- współczynnik chropowatości bezwzględnej k ,
- wysokość ciśnienia w wybranym węźle sieci ($Wys_cis_w_wezle$).

W nawiasach podano symbole atrybutów, umieszczone na schemacie drzewa decyzyjnego (rys. 1).

Wszystkie przykłady uczące, zawierające wartości podanych wyżej atrybutów, zostały opisane etykietą informującą o jego przynależności do określonej klasy. Zdefiniowano pięć klas opisujących problemy w systemie dystrybucji wody oraz jedną związaną z poprawnym układem systemu. W związku z tym drzewo decyzyjne powinno prowadzić do wyboru odpowiedniej klasy na podstawie atrybutów układu systemu dystrybucji wody. Klasy charakteryzujące stan systemu zdefiniowano w następujący sposób:

- zbyt wysokie ciśnienie w sieci (etykieta STW),
- zbyt niskie ciśnienie w sieci (etykieta STM),
- zaleca się korektę średnic w celu zmniejszenia strat ciśnienia (etykieta SR),

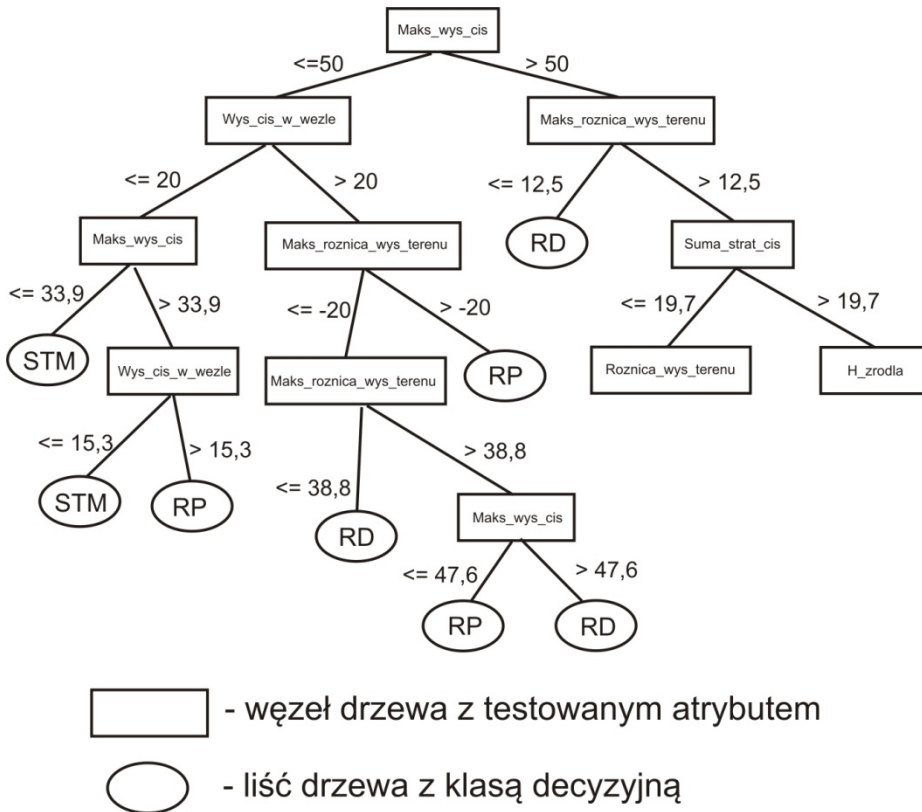
- zaleca się wyodrębnienie za danym węzłem oddzielnej strefy zasilanej przez pompownię sieciową (etykieta PO),
- zaleca się wyodrębnienie za danym węzłem oddzielnej strefy z reduktorem ciśnienia (etykieta RD),
- prawidłowy przebieg linii ciśnień (etykieta RP).

Utworzenie zbioru przykładów uczących wymagało opracowania procedury numerycznej, która pozwoliła automatycznie obliczać wartości atrybutów na podstawie wyników obliczeń. Przykłady uczące zapisano w postaci pliku bazy danych, w której wiersze (rekordy) bazy odpowiadają obiektom, natomiast kolumny atrybutom. Uzyskano zbiór 8427 przykładów uczących, reprezentujących wszystkie opisane powyżej klasy.

5. Indukcja drzewa decyzyjnego do oceny układu systemu dystrybucji wody

Na podstawie przygotowanego zbioru przykładów uczących dokonano indukcji drzewa decyzyjnego przy użyciu programu DeTreeX. Zgodnie z zasadą indukcji, drzewo decyzyjne może być wykorzystywane również dla obiektów opisanych atrybutami spoza zbioru uczącego, co pozwala na zastosowanie reguł zapisanych w drzewie do oceny nowych wariantów projektowych systemu dystrybucji wody.

Reguły decyzyjne uzyskane w wyniku indukcji drzewa decyzyjnego na podstawie zgromadzonych przykładów uczących pozwalają ocenić stan systemu dystrybucji wody. Przydzielenie do określonej klasy nowego przypadku spoza zbioru uczącego, otrzymanego w kolejnych obliczeniach hydraulicznych, pozwala również ustalić, jakie kroki należy podjąć, by zapewnić odpowiednie ciśnienie w poszczególnych węzłach. Program komputerowy może wskazać celowość zastosowania pompowni sieciowej, reduktora ciśnienia lub dokonania korekty średnic przewodów. Podaje również informację o ciśnieniu zbyt niskim lub zbyt wysokim oraz prawidłowym. Ze względu na wielkość otrzymanego drzewa, rysunek 1 przedstawia fragment uzyskanego drzewa decyzyjnego z atrybutami w węzłach, regułami wyboru ścieżki w drzewie oraz klasami przypisanymi do liści drzewa.



Rys. 1. Fragment drzewa decyzyjnego do oceny układu systemu dystrybucji wody

Fig. 1. Part of the decision tree for evaluation of water distribution system

5. Podsumowanie i wnioski

Zaprezentowana w niniejszym artykule metoda wnioskowania indukcyjnego za pomocą drzew decyzyjnych jest skutecznym narzędziem pozyskiwania wiedzy do systemów ekspertowych i może być również stosowana w zagadnieniach zaopatrzenia w wodę.

Zdefiniowane klasy oraz opisujące je atrybuty powinny pozwolić na ocenę układu systemu dystrybucji wody oraz podjąć ewentualne kroki mające na celu poprawę rozwiązania projektowego. Otrzymane drzewo decyzyjne pozwala wygenerować reguły decyzyjne, jako reprezentację wiedzy, które następnie mogą być wykorzystywane w systemach eksper-

towych, współpracujących z tradycyjnymi programami do obliczeń hydraulicznych sieci wodociągowych.

Systemy ekspertowe w najbliższym czasie z pewnością znajdą szerokie zastosowanie w zagadnieniach zaopatrzenia w wodę.

Literatura

1. **Bolc, L., Zaremba, J.:** *Wprowadzenie do uczenia się maszyn*. Akademicka Oficyna Wydawnicza RM, Warszawa 1992.
2. **Bubnicki Z., Grzech A.:** *Inżynieria wiedzy i systemy ekspertowe*. Oficyna Wydawnicza Politechniki Wrocławskiej. Wrocław 2000.
3. **Chromiec J., Strzemieczna E.:** *Sztuczna inteligencja. Metody konstrukcji i analizy systemów eksperckich*. Akademicka Oficyna Wydawnicza PLJ. Warszawa 1994.
4. **Cichosz P.:** *Systemy uczące się*. WNT. Warszawa 2000.
5. **Epp R., Fowler A. G.:** *Efficient Code for steady state Flows in Networks*. Journal of the Hydraulics Division, Proceedings of the American Society of Civil Engineers, Vol. 96, No. HY1, 43–56 (1970).
6. **Knapik K.:** *Dynamiczne modele w badaniach sieci wodociągowych*. Wydawnictwo Politechniki Krakowskiej. Kraków 2000.
7. **Koelle, E.:** *Advanced Water Distribution Modeling And Management*. Haestad Methods Solution Center, Haestad Press, 2003.
8. **Konar A.:** *Computational Intelligence: Principles, Techniques and Applications*. Springer-Verlag 2005.
9. **Kubat M., Bratko I., Michalski R.S.:** *Review of machine learning methods, learning and data mining*. John Wiley & Sons, 3–70 (1998).
10. **Mulawka J.J.:** *Systemy ekspertowe*. WNT, Warszawa 1996.
11. **Michalik K.:** *PC-Shell 4.1 dla Windows 9x/NT/2000. Szkieletowy system ekspertowy. Część 1. Podręcznik użytkownika*. AITECH, Katowice 2004.
12. **Michalik K.:** *PC-Shell 4.1 dla Windows 9x/NT/2000. Szkieletowy system ekspertowy. Część 2. Podręcznik inżyniera wiedzy*. AITECH, Katowice 2004.
13. **Michalik K.:** *DeTreex 4.0 dla Windows 9x/NT/2000. Indukcyjny system pozyskiwania wiedzy. Podręcznik użytkownika*. AITECH, Katowice 2003.
14. **Niederliński A.:** *Regulowe systemy ekspertowe*. Wydawnictwo Pracowni Komputerowej Jacka Skalmierskiego. Gliwice 2000.

15. **Ormsbee L.E.:** *The History of Water Distribution Network Analysis: The Computer Age*. Proceedings of the 8th Annual Water Distribution Systems Analysis Symposium, ASCE, 1–6 (2006).
16. **Quinlan J.R.:** *Induction of decision trees*. Machine Learning. Vol. 1, 81–106 (1986).
17. **Quinlan J.R.:** *C 4.5 Program for Machine Learning*. Morgan Kaufmann, San Mateo, 1993.
18. **Quinlan J.R.:** *Improved use of continuous attributes in C4.5*. Journal of Artificial Intelligence Research, Vol. 4, 77–90 (1996).
19. **Rossman, L.A.:** *EPANET 2 User's manual, EPA/600/R-00/057*. National Risk Management Research Laboratory, U.S. Environmental Protection Agency, Cincinnati, OH, 2000.
20. **Rutkowski L.:** *Metody i techniki sztucznej inteligencji. Inteligencja obliczeniowa*. Wydawnictwo Naukowe PAN. Warszawa, 2009.
21. **Shamsi U.M.:** *GIS applications for water, wastewater, and stormwater systems*. Taylor & Francis, 2005.
22. **Taher S.A., Labadie, J.W.:** *Optimal design of water-distribution networks with GIS*. Journal of Water Resources Planning and Management, ASCE, Vol. 122, No.4, 301–311 (1996).
23. **Walski, M.T., Chase, D.V., Savic, D.A., Grayman, W.M., Beckwith S., Koelle E.:** *Advanced Water Distribution Modeling and Management*. Haestad Press, Waterbury, CT, 204–207 (2003).

Expert System for Evaluation of Water Distribution System Created with an Inductive Inference

Abstract

At present often in computer programs the methods of computational intelligence are used, in this expert systems. In building of expert systems the process of knowledge acquisition is one of the principle problem. The improvement of knowledge acquisition is received by use of the machine learning methods. The most popular strategy of knowledge acquisition in the machine learning methods is the inductive inference, in this induction of decision trees. Inductive inference is the process of reaching a general conclusion from specific examples. This paper presents results of the induction of the decision tree intended to evaluation of water distribution system. Rules kept in the decision tree make possible to estimate the new project variants of the water supply network. Applying the induction of decision trees entails the preparation of a set of examples learners. Collection of examples should be representative and sufficient-

ly describe the specific features of the problem. Using information on the eight water supply systems and the parameters of their work, the computations were performed water distribution system. The calculation results were the basis for the calculation defined in this study the variables that characterize the solution of the water distribution system. Five classes are defined to describe the water supply problems due to improper water distribution system and one associated with the correct solution. Each class will be selected on the basis of defined variables. Therefore, the decision tree should lead to the assignment of variables describing the system of water distribution to the appropriate class, characterized by the solution of the system. Using inductive inference obtained decision rules that can be used in expert system that can work with the program for the simulation of hydraulic water distribution systems.