

Tomasz ŁUKASZEWSKI, Mariusz NOWAK

Institut Informatyki
Politechnika Poznańska

INTELIĞENTNE SYSTEMY STEROWANIA W INŻYNIERII ŚRODOWISKA Z WYKORZYSTANIEM PREDYKCJI BAZUJĄCEJ NA KLASYFIKATORZE BAYESOWSKIM

INTELLIGENT CONTROL SYSTEMS IN ENVIRONMENTAL
ENGINEERING WITH THE PREDICTION BASED
ON THE BAYESIAN CLASSIFIER

In this paper we present intelligent control systems with the prediction of the water consumption. The goal of the use of such a prediction is to achieve a very high quality of control systems with reduced cost of control processes. The prediction allows to minimize or even eliminate the changes of control systems comparing to the classical approach without the prediction. The novelty of our approach is to take into account imprecise values in the prediction process. We show, that imprecise values, that were a problem in classical approaches, can be used in order to improve the prediction accuracy. The ontological Bayesian classifier, that is used for the prediction, is a new approach in the machine learning, and is able to maintain imprecise data.

1. Wprowadzenie

Transport cieczy pochłania 20-30% całkowitego zużycia energii w gospodarce. Szacuje się, że w Polsce na wszelkiego rodzaju transport cieczy przypada 30 mld kWh, co jest jednoznaczne z emisją 35 mln ton CO₂ rocznie [1, 4]. Celowe jest zatem poszukiwanie metod zmniejszania energochłonności procesów transportu cieczy poprzez zwiększanie sprawności instalacji pompowych.

Jednym z miejsc, gdzie instalacje pompowe generują największy odsetek całociesięwych kosztów funkcjonowania są instalacje wodociągowe. Pobór wody (zużycie, rozbiór) i zapotrzebowanie na nią przez konsumentów ma charakter deterministyczno-losowy. Wielkość zużycia wody przez odbiorców podlega bowiem pewnym prawidłowościom obserwowanym o odpowiednich godzinach, dniach tygodnia, pór roku czy też typu odbiorcy. Na te prawidłowości w poborze wody dodatkowo mogą nakładać się różnego rodzaju czynniki losowe wynikające z nagłej zmiany pogody, imprez sportowych czy też awarii.

W klasycznym podejściu (bez prognozowania poboru wody) sterowanie realizowane jest w oparciu o bieżącą analizę różnicy pomiędzy zapotrzebowaniem na dany zasób,

a dysponowaną jego ilością. Uzyskanie wysokiej jakości sterowania wymaga ponoszenia wysokich kosztów związanych np. z częstą ingerencją w sposób pracy urządzeń wykonawczych.

Jednym z podejść uzyskania wysokiej jakości sterowania przy niższych kosztach jest zastosowanie prognozowania poboru wody w celu zmniejszenia stopnia ingerencji w sposób pracy urządzeń. Przykładowo, mając dostęp do prognozowanego poboru wody oraz ustalając możliwość wystąpienia dopuszczalnej (bezpiecznej) różnicy pomiędzy poborem wody a dysponowaną jej ilością, można zmniejszyć stopień ingerencji w sposób pracy urządzeń wykonawczych a nawet nie dokonywać żadnej ingerencji, w sytuacjach gdy klasyczne sterowanie (bez prognozowania poboru wody) wymagało by znacznie większego stopnia ingerencji.

Zmniejszony stopień ingerencji można zastosować w sytuacji przewidywanej zmiany poboru wody, która mogłaby spowodować przekroczenie dopuszczalnej różnicy i konieczność ingerencji w sposób pracy urządzeń w sposób dość znaczny (np. zwiększony pobór wody w godzinach wieczornych czy też godzinach rozpoczęcia prac fabryk itp.). W takiej sytuacji, zastosowanie odpowiednio wcześniej zmniejszonego stopnia ingerencji, w celu dostosowania dysponowanej ilości wody do przewidywanej zmiany poboru wody (w naszym przykładzie byłoby to stopniowe zwiększanie dysponowanej ilości wody przed przewidywanym momentem zwiększonego poboru), powinno być mniej kosztowne niż opóźniona ale znaczna ingerencja w sposób pracy urządzeń wykonawczych.

Brak ingerencji można zastosować w sytuacji wystąpienia dopuszczalnej różnicy pomiędzy poborem wody a dysponowaną jej ilością, gdy przewidywane jest pozostanie tej różnicy w dopuszczalnych ustalonych granicach (np. godziny nocne, godziny pracy fabryk i zakładów). W takiej sytuacji brak ingerencji nie generuje żadnych dodatkowych kosztów w porównaniu do jakiegokolwiek ingerencji w sposób pracy urządzeń wykonawczych.

Prognozowanie poboru wody można realizować na wiele sposobów. Najmniej kosztownym rozwiązaniem jest posiadanie pełnych danych dla każdej godziny, dnia tygodnia i pory roku [3]. Posiadanie pełnych danych wymaga jednak poświęcenia pewnego czasu na ich zebranie. Ponadto, posiadanie pełnych danych byłoby możliwe, gdyby zostały uwzględnione takie czynniki jak temperatura, wilgotność, opady, imprezy masowe, awarie, anomalie pogodowe. Zauważmy, że w początkowej fazie pracy systemu sterowania oraz zakładając, że odbiorcy wykazują pewną dynamikę zmian, nie można mówić o możliwości posiadania pełnych danych. Podsumowując, zakładamy że nie jest możliwe zebranie pełnych danych, z tego powodu prognozowanie poboru wody musi być zrealizowane w inny sposób, z danych częściowych.

Aktualnie stosowanym w badaniach naukowych podejściem do problemu generalizacji danych (w naszym wypadku prognozowania poboru wody z danych częściowych) jest uczenie maszynowe, a dokładniej mówiąc uczenie nadzorowane i klasyfikatory. Ze względu na charakter problemu, w którym mogą wystąpić wartości nieprecyzyjne, proponujemy w pracy zastosowanie ontologicznego klasyfikatora bayesowskiego, który jest uogólnieniem klasyfikatora bayesowskiego w celu uwzględnienia tego typu nieprecyzyjnych danych [2].

2. Problem sterowania w instalacjach pompowych

Proces projektowania lub modernizacji instalacji pompowej wymaga określenia optymalnego układu rurociągów oraz optymalnej liczby możliwie najlepszych pomp. Najlepsze pompy charakteryzują się racjonalnymi znamionowymi parametrami pracy, najbardziej ekonomicznym sposobem regulacji oraz najmniejszym zużyciem energii elektrycznej. Wybór sposobu regulacji polega na ustaleniu jaka liczba pomp w zespole ma posiadać regulowaną prędkość obrotową, czy i w jakim zakresie wydajności ma być stosowana regulacja dławieniowa oraz jaka liczba pomp będzie pracowała jednocześnie. Istotnym jest także sposób zmian prędkości obrotowej pompy oraz uwzględnianie wpływu prędkości obrotowej na wartość sprawności pompy lub zespołu pomp [1].

Celem regulacji w instalacjach pompowych jest dostosowanie parametrów pracy pompy, takich jak wydajność lub wysokość podnoszenia, do zmieniających się w czasie wymagań odbiorcy cieczy. W zależności od rodzaju pompy stosuje się regulacje: dławieniową, zmiennoodrotową lub polegającą na zmianie kąta ustawienia łopatek wirnika. Skokową zmianę wydajności systemu, w którym pracuje więcej niż jedna pompa uzyskuje się poprzez włączanie lub wyłączanie odpowiedniej liczby pomp.

W pompowniach wodociągowych wydajność instalacji zależy od zmiennego zapotrzebowania, którego nie da się z góry dokładnie ustalić. Posiadając dostęp do danych statystycznych, można dokonać prognozowania zakresu zmienności zapotrzebowania na wodę w ciągu doby, miesiąca lub roku.

Problem optymalnego sterowania w dużych instalacjach wodociągowych nie polega wyłącznie na minimalizacji zużycia energii w ciągu doby przez silniki napędowe zainstalowanych pomp, pracujących w różnych przedziałach czasu i z różnymi wydajnościami. Należy w taki sposób sterować pracą pomp, aby całkowite chwilowe zużycie energii zużywanej przez zainstalowane pompy było minimalne. Dla tak zdefiniowanej funkcji celu, w sytuacji odpowiednio zmniejszonego zapotrzebowania na wodę, korzystniejsze jest wyłączenie niektórych pomp niż praca wszystkich pomp z obniżonymi prędkościami obrotowymi. Inne możliwe do zastosowania algorytmy bieżącej regulacji wydajności obiektów pompowych pod kątem optymalizacji energetycznej uwzględniają zmiany sprawności w odpowiednich punktach pracy zespołów pomp, sugerując pracę z najwyższymi ich sprawnościami osiąganymi przy różnych prędkościach obrotowych.

Praktyczne wykorzystanie metod optymalizacji pracy pomp poprzez zastosowanie odpowiednich algorytmów sterowania jest możliwe pod warunkiem wykorzystania komputerowych systemów sterowania. Konieczne jest posiadanie odpowiednich narzędzi symulacyjnych w celu doboru odpowiedniej strategii sterowania, a docelowo wdrożenie globalnej optymalizacji pracy kilku połączonych zespołów pompowych pracujących w rozbudowanym systemie wodociągowym.

W tej pracy nie rozważamy problemu sterowania w instalacjach pompowych ze względu na fakt, że jest to problem niezależny od problemu prognozowania poboru wody. Jakkolwiek dalsze badania mogą dotyczyć problemu sterowania, a nawet połączenia rozwiązań dla obu wyżej wymienionych problemów.

3. Prognozowanie poboru wody z wykorzystaniem ontologicznego klasyfikatora bayesowskiego

We wstępie do pracy zauważyliśmy, że jednym podejściem uzyskania wysokiej jakości sterowania przy niższych kosztach jest zastosowanie prognozowania poboru wody w celu zmniejszenia stopnia ingerencji w sposób pracy urządzeń. Ze względu na charakter problemu, w którym mogą wystąpić wartości nieprecyzyjne, proponujemy w pracy zastosowanie ontologicznego klasyfikatora bayesowskiego, który jest uogólnieniem klasyfikatora bayesowskiego w celu uwzględnienia danych nieprecyzyjnych. W dalszej części przedstawiony zostanie klasyfikator bayesowski, oraz ontologiczny klasyfikator bayesowski.

3.1. Klasyfikator bayesowski

Naiwny klasyfikator bayesowski, pomimo prostoty i silnych założeń o niezależności warunkowej, działa bardzo dobrze, konkurencyjnie z innymi dobrze znanymi technikami indukcyjnymi takimi jak: drzewa decyzyjne i sieci neuronowe. Naiwny klasyfikator bayesowski jest często używany do problemów klasyfikacji, w których próbuje się zbudować klasyfikator z danego zbioru T przykładów uczących z etykietami klasy. Przyjmuje się, że dany jest zbiór n atrybutów A_1, A_2, \dots, A_n . Przykład (uczący lub testowy) jest reprezentowany przez wektor (v_1, v_2, \dots, v_n) , gdzie v_i jest precyzyjną wartością atrybutu A_i . C reprezentuje klasę a C_j reprezentuje wartość, jaką przyjmuje klasa (etykieta klasy). Klasyfikator bayesowski (również naiwny klasyfikator bayesowski) klasyfikuje nową obserwację E przez wybór klasy C_j o największym prawdopodobieństwie $P(C_j|E)$, tak jak przedstawiono poniżej:

$$P(C_j|E) = \frac{P(C_j)P(E|C_j)}{P(E)} \quad (1)$$

$P(E)$ jest ignorowane w przypadku, gdyż jest takie samo dla każdej klasy i nie wpływa na wartości relatywne tych prawdopodobieństw:

$$P(C_j|E) \propto P(C_j)P(E|C_j) \quad (2)$$

Ponieważ E jest kompozycją n wartości, możemy rozbić to wyrażenie:

$$P(C_j|v_1, v_2, \dots, v_n) \propto P(C_j)P(v_1, v_2, \dots, v_n|C_j) \quad (3)$$

Gdzie $P(v_1, v_2, \dots, v_n|C_j)$ jest prawdopodobieństwem wystąpienia przykładu E pod warunkiem przynależności do klasy C_j ; $P(C_j)$ jest prawdopodobieństwem a priori klasy C_j .

Naiwny klasyfikator bayesowski zakłada, że atrybuty są niezależne warunkowo od podanych wartości klasy, co daje:

$$P(C_j|v_1, v_2, \dots, v_n) \propto P(C_j)\prod P(v_i|C_j) \quad (4)$$

$P(v_i|C_j)$ jest prawdopodobieństwem, że instancja klasy C_j ma wartość atrybutu A_i równą v_i . Prawdopodobieństwa w powyższym wzorze muszą być oszacowane z przykładów uczących, na przykład używając częstotliwości względnej:

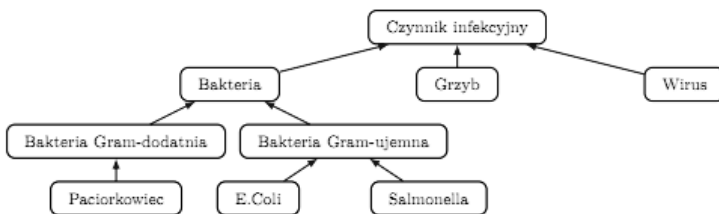
$$P(C_j) = \frac{n_j}{n} \quad P(v_i|C_j) = \frac{n_{ij}}{n_j} \quad (5)$$

Gdzie n jest liczbą przykładów uczących, n_j jest liczbą przykładów uczących z etykietą klasy C_j , n_{ij} jest liczbą przykładów uczących, w których atrybut A_i ma wartość v_i i etykietą klasy jest C_j .

3.2. Ontologia wartości atrybutu

We wprowadzeniu wskazaliśmy, że w rozważanym problemie prognozowania poboru wody mogą wystąpić wartości nieprecyzyjne. Wartości te mogą być uwzględnione przez klasyfikator przez użycie języka reprezentacji wiedzy, który umożliwi modelowanie różnych poziomów precyzji opisu. Poziomy te powinny odzwierciedlać wiedzę dziedzinową i nie powinny być tworzone w sposób dowolny. W pewnych dziedzinach mogą być obserwowane relacje hierarchiczne między pojęciami, które mogą być wykorzystane w języku opisu przykładów. Taka wiedza jest często dostępna w formie ontologii. W ten sposób, precyzyjne i nieprecyzyjne wartości atrybutu A mogą być reprezentowane przez strukturę hierarchiczną, nazwaną ontologią wartości atrybutu (ang. *attribute value ontology*) (AVO). Definicja ontologii wartości atrybutu jest podana w pracy [2].

Przyjmujemy, że w takiej ontologii precyzyjne wartości atrybutu są reprezentowane przez pojęcia prymitywne, a nieprecyzyjne, niepełne i nieznanne wartości przez pojęcia abstrakcyjne. Przed zdefiniowaniem ontologii wartości atrybutu (AVO), pokazany zostanie przykład demonstrujący hierarchię pojęć.



Rys. 1. Przykład ontologii wartości atrybutu
Fig. 1. Example of an attribute value ontology (AVO)

Przykładowo, rozważmy następujący problem medyczny. W celu określenia poprawnego leczenia, należy określić czynnik, który spowodował infekcję. Poszczególne wartości atrybutu (rozważamy tylko jeden atrybut) są następujące: Paciorkowiec, E.Coli, Salmonella, Grzyb, Wirus. AVO opisującą tę dziedzinę czynników infekcyjnych przedstawiono na rysunku 1. Pojęcia prymitywne są następujące: Paciorkowiec, E.Coli, Salmonella, Grzyb, Wirus. Natomiast pojęcia abstrakcyjne są następujące: Czynnik infekcyjny, Bakteria, Bakteria Gram-dodatnia, Bakteria Gram-ujemna.

Należy zauważyć, że Paciorkowiec nie jest jedyną Bakterią Gram-dodatnią w rzeczywistym świecie. Niestety nasza hierarchia z pewnych względów nie zawiera pojęć reprezentujących inne Bakterie Gram-dodatnie. Dlatego, pojęcie Bakteria Gram-dodatnia prawidłowo powinno być interpretowane jako: Paciorkowiec lub inna Bakteria Gram-dodatnia. Podobnie, pojęcie Bakteria Gram-ujemna powinno być interpretowane jako: E.Coli, Salmonella lub inna Bakteria Gram-ujemna. Interpretacja ta wynika z przyjętego w tym przykładzie założenia o otwartości świata (ang. *open world assumption*).

3.3. Ontologiczny klasyfikator bayesowski

W proponowanym podejściu ontologicznym, naiwny klasyfikator bayesowski wymaga rozszerzenia w celu oszacowania $P(c_i|C_j)$, gdzie c_i jest pojęciem prymitywnym lub abstrakcyjnym ontologii wartości atrybutu A . Zaproponowane podejście jest uogólnieniem podejścia klasycznego. W klasycznym podejściu, wartości atrybutów mogą być interpretowane jako pojedynczy poziom precyzji i nowy przykład jest klasyfikowany przez użycie przykładów uczących opisany tą samą wartością atrybutu. Jednakże, w sytuacji gdy klasyfikowany przykład jest opisany pojęciem abstrakcyjnym, do jego klasyfikacji należy rozważyć użycie przykładów uczących opisanych:

- tym samym pojęciem abstrakcyjnym,
- podpojęciami tego pojęcia abstrakcyjnego.

Przypomnijmy, że dla danego pojęcia c_i w AVO, każde pojęcie które jest bardziej precyzyjne niż c_i , jest potomkiem pojęcia c_i . Zatem, w celu oszacowania prawdopodobieństwa $P(c_i|C_j)$, na przykład używając częstotliwości względnej, użyjemy tej własności i otrzymamy:

$$P(c_i|C_j) = \frac{n_{ij} + \sum_{c_k \in desc(c_i)} n_{kj}}{n_j} \quad (6)$$

Gdzie n_j jest liczbą przykładów uczących z etykietą klasy C_j , n_{ij} jest liczbą przykładów uczących, w których atrybut ma wartość c_i i etykietą klasy jest C_j , n_{kj} jest liczbą przykładów uczących, w których wartością atrybutu A jest c_k i etykietą klasy jest C_j , $desc(c_i)$ jest zbiorem pojęć, które są potomkami pojęcia c_i . Szczegółowy opis ontologicznego klasyfikatora bayesowskiego jest podany w pracy [2].

3.4. Zastosowanie klasyfikatora do prognozowania poboru wody

W celu zastosowania klasyfikatora do prognozowania poboru wody należy przede wszystkim ustalić, jakiego rodzaju atrybuty będą brane pod uwagę i jakie będą ich wartości. Ze względu na specyfikę problemu, gdzie wielkość zużycia wody przez odbiorców podlega pewnym prawidłowościom obserwowanym w odpowiednich godzinach, dniach tygodnia, pór roku oraz jest zależna od typu odbiorcy, proponujemy użycie następujących atrybutów:

- godzina (zbiór wartości to pełne godziny),
- dzień tygodnia (roboczy, sobota, niedziela, święto),
- pora roku (wiosna, lato, jesień zima).

Ze względu na fakt, że konieczne może być oszacowanie wielkości zużycia wody zanim dokonamy zebrania danych historycznych, można dokonać tego dla każdego typu odbiorcy oddzielnie (firmy, użytkownicy prywatni, działkowicze itp.), a następnie wykonać agregację tych szacunkowych danych. Po zebraniu rzeczywistych danych, dane szacunkowe mogą pozostać w zbiorze danych uczących lub można je usunąć po pewnym czasie.

Ze względu na zakładaną dynamikę wielkości zużycia wody, zakładamy że będziemy pamiętać wartości zużycia z kilku ostatnich lat.

Głównym celem tej pracy jest jednakże zastosowanie ontologicznego klasyfikatora bayesowskiego do prognozowania poboru wody. Klasyfikator ten pozwala na użycie wartości nieprecyzyjnych. Przykładowo, w okresie przejściowym pomiędzy wiosną a latem, ontologiczny klasyfikator bayesowski pozwala użyć danych uczących (szacunkowych, historycznych) zarówno z okresu wiosny jak i lata, co jest bardzo intuicyjnym podejściem, a nie było to możliwe w klasycznych podejściach, operujących na danych precyzyjnych. Tak samo, w przypadku trudności określenia jak zachowują się odbiorcy w dniu tygodnia, który jest dla części osób dniem roboczym a dla innych wolnym od pracy, możemy użyć do prognozowania danych uczących dla obu rodzaju dni tygodnia, co ponownie nie było dotychczas możliwe w klasycznych podejściach. Są to jedynie przykładowe możliwości użycia wartości nieprecyzyjnych na drodze łączenia wartości precyzyjnych, reprezentowane przez pojęcia abstrakcyjne AVO. Dokładny dobór wartości nieprecyzyjnych wymaga uwzględnienia specyfiki danej konfiguracji odbiorców oraz przeprowadzenia eksperymentów obliczeniowych. Problematyką tą zajmujemy się w dalszych badaniach nad inteligentnymi systemami sterowania.

4. Podsumowanie

Inteligentne systemy sterowania okazują się konieczne praktycznie w każdym złożonym systemie technologicznym. Liczba kryteriów oceny skuteczności tego typu systemów sterowania jest różna, jednak w znacznej większości przypadków jest to problem wielokryterialny. Dlatego tak ważne jest efektywne działanie tych systemów pod każdym względem. Jednym z kryteriów jest wysoka jakość sterowania. W pracy wskazaliśmy, że uzyskanie wysokiej jakości sterowania w systemach dystrybucji wody wymaga ponoszenia wysokich kosztów związanych np. z częstą ingerencją w sposób pracy urządzeń wykonawczych. Jednym z podejść uzyskania wysokiej jakości sterowania przy niższych kosztach jest zastosowanie prognozowania poboru wody w celu zmniejszenia stopnia ingerencji w sposób pracy urządzeń [3]. Wykazaliśmy teoretycznie, że użycie wartości nieprecyzyjnych pozwala uzyskać wyższą trafność prognozowania w stosunku do podejścia klasycznego. Zaproponowane w pracy podejście ontologiczne jest podejściem nowatorskim, ponieważ pozwala użyć wartości nieprecyzyjnych, które dotychczas stanowiły poważny problem w metodach klasycznych. Dalsze prace badawcze przewidują weryfikację skuteczności działania takiego klasyfikatora.

Bibliografia

- [1] Jędrał W., *Efektywność energetyczna pomp i instalacji pompowych*, Krajowa Agencja Poszanowania Energii S.A., Warszawa 2007
- [2] Łukaszewski T., Józefowska J., Ławrynowicz A., Józefowski Ł., *Handling the description noise using an attribute value ontology*, Control and Cybernetics, 40 (2), 275-292, 2011
- [3] Siwoń Z., Cieżak W., Cieżak J., *Wybrane modele neuronowe dobowych profili godzinowego rozbioru wody*, Zaopatrzenie w wodę, jakość i ochrona wód – zagadnienia współczesne, J. F. Lemański, S. Zabawa (red.), Wyd. PZiTS O/Wielkopolski, Poznań, Polska 2010, (99-114 t. II)
- [4] Rozporządzenie ministra infrastruktury z dnia 14 stycznia 2002r. w sprawie określenia przeciętnych norm zużycia wody (Dz. U. 2002 nr 8 poz. 70)