

Pozyskiwanie wiedzy z danych katalogowych komponentów systemów pomiarowych

Jan Jagielski
Izabela Skorupska

W artykule opisano sposób automatycznej generacji bazy wiedzy z danych katalogowych przetworników. Dane katalogowe jako baza danych pozwoliły wygenerować drzewo decyzyjne. Na podstawie wiedzy zawartej w drzewie decyzyjnym, za pomocą pakietu Sphinx, zbudowano bazę wiedzy dla systemu eksperckiego. Stworzona aplikacja jest łatwa w użyciu i za jej pomocą w prosty sposób można wspomóc podjęcie decyzji dotyczącej wyboru czujnika pomiarowego.

Knowledge Discovery in Databases from catalogue cards of temperature sensors

Create automatically knowledge base from catalogue cards of temperature sensors are presented in this paper. Catalogue data as database may be used to generate decision tree. Decision tree is the most understandable form of knowledge base to human beings. We may create automatically knowledge base from decision tree. This knowledge base may be used to build expert system. They generate automatically knowledge base from database. Created application is easy to use. Methods and computer program (e.g. Sphinx) are described in this paper

1. Wprowadzenie

System pomiarowy jest zbiorem elementów realizujących specyficzne funkcje i objętych wspólnym sterowaniem. Umożliwia on automatyczne wykonywanie pomiarów różnych wielkości w wielu punktach badanego obiektu. Odpowiednio przetworzone dane katalogowe elementów tego systemu mogą tworzyć bazę danych, która może być źródłem wiedzy dla projektanta pozyskanej metodą eksploracji danych.

2. Możliwości i wymagania metod eksploracji danych

Eksploracja danych (spotyka się również określenie drążenie danych) (*Data Mining*) to jeden z etapów procesu odkrywania wiedzy z baz danych (*Knowledge Discovery in Databases* – KDD) jako odkrywanie obowiązujących, a dotychczas nieznanych potencjalnie użytecznych i zrozumiałych wzorców w zbiorach danych. Metodami szczegółowymi eksploracji danych są: klasyfikacja, grupowanie, odkrywanie wzorców sekwencji i odkrywanie reguł asocjacyjnych. Ta ostatnia metoda obejmuje znajdowanie związków pomiędzy występowaniem grup atrybutów elementów lub ich wartościami. Eksploracja danych bazuje na takich naukach, jak: statystyka (statystyczna analiza wielowymiarowa) i uczenie maszynowe. Praktyka eksploracji danych polega na wykorzystaniu odpowiednich programów komputerowych do znajdowania ukrytych, wcześniej nieznanych zależności z dostępnych danych [1, 2, 10]. Niektóre z nich (z możliwością zastosowania w biznesie) zostały przedstawione w [3]. Zastosowanie metod eksploracji danych najczęściej obejmuje te miejsca, w których stosuje się systemy informatyczne,

Jan Jagielski – e-mail: J.Jagielski@ime.uz.zgora.pl
Uniwersytet Zielonogórski, Instytut Metrologii Elektrycznej;
Izabela Skorupska – e-mail: I.Skorupska@weit.uz.zgora.pl
Studium Doktoranckie na Wydziale Elektrotechniki, Informatyki i Telekomunikacji Uniwersytetu Zielonogórskiego

m.in. w celu gromadzenia i pozyskanych danych w postaci baz danych. Wszędzie tam, gdzie istnieje już baza danych, pojawia się możliwość ich analizy w celu odkrycia nieznanego dotąd wiedzy. Dziedziny, w których szeroko stosuje się eksplorację danych to: technika, medycyna, diagnostyka oraz biznes. Możliwość taka zostanie przedstawiona w artykule na przykładzie procedury budowy systemu doradczego, który na podstawie danych katalogowych będzie mógł wspomagać projektanta w wyborze elementów składowych systemu pomiarowego.

3. Przekształcenie danych katalogowych w bazy danych

Dane katalogowe [5, 6, 7, 8, 9] muszą być przekształcone w bazy danych zgodnie z wymaganiami stosowanego programu eksploracji danych. Przeanalizowane zostały karty katalogowe przetworników temperatury. Na ich podstawie zdecydowano się zbudować bazę danych, która po przetworzeniu za pomocą pakietu SPHINX firmy Aitech stanie się źródłem wiedzy dla systemu eksperckiego. W skład pakietu SPHINX wchodzi m.in. program DeTreex, który umożliwia generowanie drzew decyzyjnych na podstawie plików baz danych oraz program PC Shell umożliwiający budowę systemów eksperckich.

Budowa takiego systemu składa się z kilku etapów:

1. Analiza kart katalogowych komponentów systemu pomiarowego
2. Określenie dziedziny badanego problemu
3. Wygenerowanie drzewa decyzyjnego
4. Stworzenie bazy danych komponentów zgodnie z wymaganiami systemu DeTreex
5. Automatyczne wygenerowanie bazy wiedzy z drzewa decyzyjnego
6. System decyzyjny z bazą wiedzy.

Tab. 1. Fragment pliku uczącego programu DeTreex
Tab. 1. Fragment of file of DeTreex program

ATRYBUT/CZUJNIK	PTR-21	APAQ-HCF	APAQ-LC
Zakres temperatury otoczenia – wartość górna	55	85	75
Zakres temperatury otoczenia – wartość dolna	-25	-40	-20
Wilgotność względna	90	95	95
#Wielkość wyjściowa	prąd	napięcie	napięcie
Zakres zmian wielkości wyjściowej – wartość dolna	4	0	0
Zakres zmian wielkości wyjściowej – wartość górna	20	50	50
Stała czasowa	1	0,2	0,2
Błąd podstawowy	0,1	0,2	0,2
#Decyzja	zastosować	zastosować	nie zastosować

3.1. Określenie dziedziny badanego problemu

Dziedzinę rozpatrywanego problemu określa się po szczegółowej analizie kart katalogowych przetworników:

Temperatura otoczenia

– *wartość dolna*: liczba rzeczywista z przedziału <-50, 2000>

– *wartość górna*: liczba rzeczywista z przedziału <-50, 2000>

Wilgotność względna: liczba całkowita z przedziału <0, 100>

Pozycja pracy: dowolna, pozioma, pionowa

Czas nagrzewania: liczba rzeczywista z przedziału <0, 60>

Typ czujnika: temperatury

Zakres pomiarowy

– *wartość dolna*: liczba całkowita

– *wartość górna*: liczba całkowita

Wielkość wyjściowa: prąd, napięcie

Zakres zmian wielkości wyjściowej

– *wartość dolna*: liczba całkowita

– *wartość górna*: liczba całkowita

Stała czasowa: liczba całkowita

Błąd podstawowy: liczba całkowita

Sposób montażu: listwa, naścienny, kanał

Masa: wartości – liczby całkowite

Cena: wartość

Klasyfikowany obiekt (problem):

Decyzja: zastosować, nie zastosować

Po ustaleniu dziedziny problemu należy stworzyć bazę danych, z niej plik uczący i testujący (zgodnie z wymaganiami programu DeTreex) oraz poddać go dalszym etapom przekształceń.

3.2. Generowanie drzewa decyzyjnego

Po ustaleniu dziedziny problemu należy stworzyć plik uczący i testujący zgodnie z wymaganiami programu DeTreex. Przykładowe rekordy pliku uczącego zostały przedstawione w tabeli 1.

Ze względu na rozmiar tabeli w artykule ograniczono się do zaprezentowania wybranych atrybutów czujników pomiarowych oraz zamieniono miejscami wiersze z kolumnami.

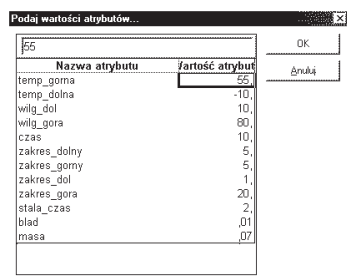
Dużym ułatwieniem dla użytkowników programu jest automatyzacja wielu procesów, w tym budowy drzewa za pomocą opcji *Generuj drzewo*. W rozpatrywanym przypadku zostało wygenerowane drzewo decyzyjne ilustrujące problem wyboru przetwornika temperatury.

W celu zweryfikowania poprawności przygotowanego drzewa decyzyjnego przeprowadzono test. W przypadku programu DeTreeX 4.0 polegał on na przygotowaniu pliku testowego zawierającego przykładowe dane sprawdzające. Po przeprowadzeniu prób okazało się, że drzewo jest poprawne i można przystąpić do budowy bazy wiedzy.

3.3. Generowanie systemu eksperckiego

Stworzone drzewo pozwala automatycznie wygenerować bazę wiedzy dla programu PC Shell i w dalszej kolejności stworzyć aplikację, która będzie wspomagała projektanta systemów pomiarowych w doborze czujnika temperatury i w wersji rozwiniętej innych komponentów składowych systemu pomiarowego.

Na rys. 1 przedstawiono ekran stworzonej aplikacji, w którym należy podać wartości poszczególnych atrybutów charakteryzujących przetwornik tempe-



Rys. 1. Ekran programu PC Shell (wybór wartości atrybutów)
Fig. 1. Screen of PC Shell program (choose value of attribute)

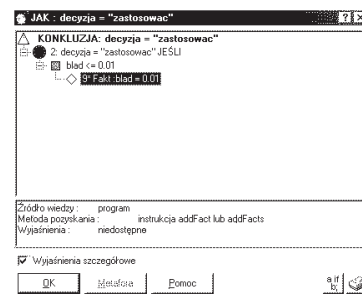
ratury.

Po akceptacji wprowadzonych atrybutów program (w zależności od podanych wielkości) może zadać dodatkowe pytania dotyczące wybieranego czujnika. Końcowym efektem działania programu jest decyzja podpowiadająca, czy w systemie pomiarowym zastosować czujnik temperatury o podanych parametrach.

4. Podsumowanie

Na podstawie analizy kart katalogowych stworzono bazę danych zawierającą informacje o komponentach systemu pomiarowego. Dzięki integracji elementów pakietu Sphinx 4.0 w programie DeTreeX wygenerowano drzewo decyzyjne. Wykorzystując zautomatyzowany proces konwersji, wygenerowano bazę wiedzy.

Baza ta posłużyła do stworzenia aplikacji „Wybór czujnika”, która ma za zadanie wspomagać podejmowanie decyzji dotyczącej wyboru czujnika temperatury.



Rys. 2. Ostatni ekran programu PC Shell

Fig. 2. Last screen of PC Shell program

Aplikacja ta może być rozszerzona do wersji wspomagającej projektanta przy wyborze pozostałych komponentów systemu pomiarowego.

5. Literatura

1. D. Hand, H. Mannila, P. Smyth, *Eksploracja danych*, Wydawnictwo Naukowo-Techniczne, Warszawa 2005.
2. Encyklopedia internetowa „Wikipedia”, http://pl.wikipedia.org/wiki/Eksploracja_danych.
3. J. Jagielski, I. Skorupska, *Metody pozyskiwania wiedzy z danych historycznych, Bazy danych – modele, technologie, narzędzia: architektura, metody formalne, bezpieczeństwo*, Wydawnictwo Komunikacji i Łączności, Warszawa 2005, s. 193–200.
4. J. Robin, B. Cockett, *Aspects of Expert Systems, Advanced Computing Concepts and Techniques in Control Engineering*, Ed. Michael J. Denham, Berlin, Springer-Verlag, 1988.
5. Katalog artykułów firmy APS
6. Katalog artykułów firmy ASTAT
7. Katalog artykułów firmy INTROL
8. Katalog artykułów firmy Linear Audio Research
9. Katalog artykułów firmy TEST-THERM
10. M. Muraszkiewicz, *Eksploracja danych w telekomunikacji* http://www.icie.com.pl/ARCHHIVE/dm_tel.doc
11. Pomoc programu DeTreeX
12. Pomoc programu PC Shell