

**Agnieszka Konys, Jarosław
Wątróbski, Krzysztof Małecki,
Paweł Suchy**

**Model systemu ekspertowego
ekstrakcji reguł decyzyjnych na
potrzeby wyceny nieruchomości**

Ekonomiczne Problemy Usług nr 71, 375-391

2011

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach
dozwolonego użytku.

AGNIESZKA KONYS, JAROSŁAW WĄTRÓBSKI,

KRZYSZTOF MAŁECKI, PAWEŁ SUCHY

Zachodniopomorski Uniwersytet Technologiczny

**MODEL SYSTEMU EKSPERTOWEGO
EKSTRAKЦИИ REGUŁ DECYZYJNYCH
NA POTRZEBY WYCENY NIERUCHOMOŚCI**

Wstęp

W dobie pogłębiającej się informatyzacji coraz więcej wymaga się od systemów informatycznych znajdujących zastosowanie w różnych dziedzinach życia. Jedną z nich jest rynek nieruchomości, w szczególności wycena. Jest to dziedzina, która wzbudza wiele emocji zarówno wśród osób prywatnych, jak i w małych oraz większych firmach. Zakup nieruchomości jest dla jednych i drugich nie lada wyzwaniem, a jej odpowiedni wybór, ze szczególnym uwzględnieniem racjonalnej ceny, w wielu przypadkach życiową decyzją. Również wielkie instytucje, jakimi są na przykład banki, niezwykle wagę przywiązują do poprawnej wyceny nieruchomości, mogącej być chociażby zabezpieczeniem kredytu hipotecznego, a dla inwestorów jest produktem, na którym chcą jak najwięcej zarobić. Nawet kilkuprocentowy błąd wyceny może się wiązać z relatywnie dużymi stratami ekonomicznymi.

Obecnie niewiele jest w Polsce systemów wyceny nieruchomości, które korzystałyby z masowych źródeł wiedzy na temat parametrów fizycznych i ekonomicznych nieruchomości. Proste algorytmy obliczeniowe nie

są w stanie przedstawić rzeczywistej wartości ze względu na niewielką liczbę kryteriów oraz czynników wpływających bądź mogących wpływać na wartość nieruchomości. Widać więc, że potrzebny jest system wyceny nieruchomości niwelujący ryzyko popełnienia błędu przy określaniu jej wartości, a także uwzględniający zmiany zachodzące na skutek działania różnych czynników, często trudnych do przewidzenia. System, który wykorzystuje najlepsze metody wyceny dostosowane do rynku nieruchomości i w przejrzysty sposób obrazuje wszystkie wykonane obliczenia oraz czynniki wpływające na wartość. W niniejszym artykule skoncentrowano się na badaniu modelu systemu ekspertowego wyceny nieruchomości wykorzystującego teorię zbiorów przybliżonych.

Pojęcie nieruchomości i metody wyceny

Nieruchomość jest wyodrębnionym obszarem ziemi, obejmującym także zasoby naturalne pod jej powierzchnią (lub lustrem wody), wytwory przyrody lub produkcji trwale związane z tą powierzchnią, a także przestrzeń nad nią¹. Jak pokazuje niniejsza definicja, pojęcie nieruchomości jest bardzo rozległe i obejmuje wiele składników. Mineralia znajdujące się pod powierzchnią ziemi czy też powietrze nad nią również są uważane za część nieruchomości. Można wyróżnić kilka cech fizycznych opisujących to pojęcie: nieprzemieszczalność, trwałość, różnorodność i niepodzielność. Jednakże istotniejszym podziałem ze względu na wycenę nieruchomości są jej cechy ekonomiczne, charakteryzujące ją od strony finansowej. Są to: rzadkość, lokalizacja, współzależność, duża kapitałochłonność oraz mała płynność². Prawidłowa wycena nieruchomości jest niezbędna do przeprowadzenia transferu praw związanych z nieruchomościami. Jest również mechanizmem ograniczającym ryzyko finansowania nieruchomości czy też uzyskania zadośćuczynienia za utratę jej wartości użytkowej, zatem szacowanie wartości nieruchomości składa się z dwóch etapów³:

¹ A. Śliwiński, *Zarządzanie nieruchomościami: Podstawy wiedzy i praktyki gospodarowania nieruchomościami*, Wydanie I, Warszawa 2000.

² Wyższa Szkoła Gospodarowania Nieruchomościami w Warszawie, www.wsgn.pl.

³ D. Isaac, N. Crosby, *Metody wyceny nieruchomości*, Hongkong 1991.

- procesu wyceny – określa najbardziej prawdopodobną cenę nieruchomości na konkurencyjnym rynku w transakcji wolnorynkowej,
- procesu analizy – określa wartość transakcji po jej ukończeniu.

Metodologia wyceny nieruchomości powinna uwzględniać zarówno obecną, jak i przyszłą sytuację na rynku, a także przeciętnego nabywcę wraz z jego nieracjonalnym zachowaniem. Istniejące metody wyceny w różnym stopniu uwzględniają te czynniki, co pozwala sądzić, że żadna z nich nie daje całkowitej gwarancji poprawności otrzymanego wyniku (rzeczywistej rynkowej wartości nieruchomości), określanej na bazie zdefiniowanych podejść⁴:

- podejście porównawcze – analiza dokonanych podobnych transakcji na istniejącym rynku nieruchomości;
- podejście dochodowe – podstawą jest założenie, iż nieruchomość przyniesie zysk jej właścicielowi;
- podejście kosztowe – istotą jest założenie, że potencjalny nabywca nie będzie skłonny zapłacić więcej niż wyniosą koszty odtworzenia nieruchomości;
- podejście mieszane – wykorzystywane tylko w szczególnych przypadkach – gdy niemożliwe jest oszacowanie wartości nieruchomości poprzez zastosowanie podejścia porównawczego bądź dochodowego; nie może ono stanowić podstawy do określania wartości rynkowej nieruchomości lub wartości odtworzeniowej nieruchomości.

Systemy automatycznej wyceny nieruchomości

System ekspercki to komputerowy program, który na podstawie szczegółowej wiedzy z określonej dziedziny może podejmować decyzje i wyciągać wnioski. Wnioskowanie odbywa się tu w sposób zbliżony do rozumowania ludzkiego eksperta z danej dziedziny. Funkcje, jakie spełnia, przesądzają o jego przynależności do technik inteligentnych⁵.

Systemy do automatycznej wyceny nieruchomości są często stosowane w masowych wycenach gruntów i budynków. Ich podstawową zaletą jest duża oszczędność czasu (wynik wyceny jest otrzymywany niemal natychmiast) oraz wymierne korzyści w zakresie wykorzystanych zasobów i użytego kapitału.

⁴ E. Kucharska-Stasiak, *Nieruchomość a rynek*, Warszawa 1999.

⁵ J. Mulawka, *Systemy ekspertowe*, Warszawa, 1997.

Rozwiązanie wygenerowane przez automatyczny system jest obiektywne, a liczba parametrów podawana na wejściu przez użytkownika często ograniczona do minimum w celu wyeliminowania ludzkich błędów i uprzedzeń, które mogłyby zafałszować otrzymaną wartość nieruchomości⁶. Jednakże taki system nie jest pozbawiony wad. Największą z nich jest to, że nie uwzględnia stanu własności nieruchomości. Przyjmowana jest uśredniona wartość, która nie może w pełni odzwierciedlać rzeczywistości. Kolejnym problemem są nowe i rzadko spotykane rodzaje nieruchomości. Podobnie jak w przypadku tradycyjnych metod stosowanych przez rzeczoznawców, w zautomatyzowanym procesie wycena takich obiektów jest trudna, gdyż brak jest odniesienia do nieruchomości podobnych i nie można wykorzystać danych historycznych na temat ich sprzedaży. Dodatkowo system nie jest w stanie przy wycenie uwzględnić specyficznych aspektów dotyczących konkretnej nieruchomości (np. położenie na polu naftowym)⁷. W tabeli 1 zestawiono funkcjonujące systemy wyceny nieruchomości wraz ze zbiorem kryteriów oraz poszczególnymi ocenami.

Wymienione w tabeli systemy bazują na analizie danych dotyczących lokalnych transakcji rynkowych z niedalekiej przeszłości. Istotne jest założenie, aby wykorzystywane informacje były kompletne, spójne oraz wiarygodne. Bez złożonych mechanizmów eksploracji danych typowe systemy informacyjne oparte na logice dwuwartościowej oraz klasycznie pojmowanych zbiorach nie są w stanie rozwiązać wielu sprzeczności i niejednoznaczności, jakie występują przy przetwarzaniu rzeczywistych danych pozyskanych nie zawsze z wiarygodnych źródeł.

Kolejną istotną wadą tych systemów jest brak modułu objaśniającego, umożliwiającego zaprezentowanie czynników i zależności wpływających na uzyskanie określonego wyniku. Jedynym uzasadnieniem oszacowanej wartości badanego obiektu jest zestawienie zawartych w bazie danych informacji o historycznych zbliżonych transakcjach. Gros systemów bazuje na pojęciu, jakim jest nieruchomość podobna. Żaden z nich nie potrafił jednak wskazać przesłanek, które przyczyniły się do zaklasyfikowania danego obiektu do zbioru nieruchomości o zbliżonej charakterystyce – nieruchomości

⁶ *Automated Valuation Models*, <<http://www.appraisaltoday.com/avms.htm>> [data dostępu: 08.11.2009].

⁷ M.L. Downie, G. Robson, *Automated Valuation Models: an international perspective*, Council of Mortgage Lenders, London 2007.

podobnych, który jest bazą wiedzy w procesie estymowania przybliżonej ceny przedmiotu badań.

Tabela 1

Podsumowanie możliwości systemów wyceny

Kryterium	Zillow ^{a)}	Home-Gain ^{b)}	Value-Point4 ^{c)}	CASA ^{d)}	Snajp ^{e)}
Uwzględnianie kluczowych parametrów (metraż, rok budowy itp.)	x	x	x	x	x
Czynniki dodatkowe (uzbrojenie, basen itp.)	-	-	x	x	-
Wpływ sąsiedztwa nieruchomości na wynik	-	-	-	-	-
Wartościowanie danych pod względem stopnia ich pewności	-	-	b/d	x	-
Wykorzystanie różnych metod do procesu szacowania	-	-	x	x	-
Wycena bazująca na transakcjach historycznych	x	x	x	x	x
Rozwiązanie w postaci konkretnej wartości	x	-	x	x	x
Współczynnik wiarygodności rozwiązania	x	-	x	x	x
Zestawienie transakcji podobnych	x	x	x	x	x

^{a)} Zillow.com

^{b)} HomeGain.com

^{c)} ValuePoint®4 – First American CoreLogic, Inc., <http://www.facorelogic.com/literature/valuepoint4.jsp>

^{d)} AFR Insurance Services – Automated Valuation Models, <http://insurance.afrservices.com/Elite/contentmgmt/ContentItemView.aspx?ItemID=3dc28627-7cb3-4a1d-bdc8-60c5e82f021f>

^{e)} Snajp.pl

Źródło: opracowanie własne.

Powyższe zagadnienia nie stanowią bariery dla systemów ekspertowych opartych na teorii zbiorów przybliżonych. Zaletą podejścia zbiorów przybliżonych do analizy danych jest możliwość wytwarzania rezultatów przybliżonych. Charakteryzuje się ono dużą tolerancją na błędy oraz możliwością przetwarzania informacji niekompletnych, a także częściowo sprzecznych.

Projekt systemu ekspertowego do wyceny nieruchomości

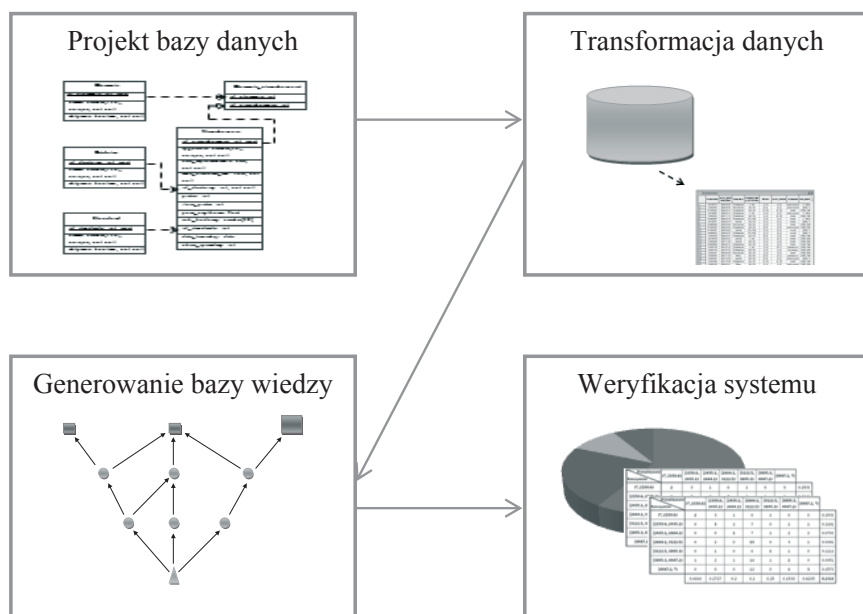
W niniejszym artykule przedstawiono założenia metodyczne oraz eksperymenty praktyczne wyceny nieruchomości z wykorzystaniem reguł decyzyjnych. W systemach eksperckich wykorzystujących aparat matematyczny teorii zbiorów przybliżonych pewne oraz możliwe związki między obiektami przedstawione są w postaci reguł decyzyjnych, stanowiących górne i dolne przybliżenia zbiorów wartości atrybutów. Taka forma zapisu wiedzy umożliwia zobrazowanie toku wnioskowania, co dodatkowo utwierdza użytkownika w słuszności podjętej decyzji.

Rozwiązania mają postać założeń dla potrzeb systemu eksperckiego bazującego na teorii zbiorów przybliżonych. Podstawowa jednostka, jaką jest zbiór przybliżony, definiowana jest następująco⁸: niech $S = (U, A, V, f)$ będzie systemem informacyjnym, gdzie U to zbiór obiektów, A – zbiór atrybutów, V jest dziedziną atrybutów, a f funkcją informacji. Zbiory przybliżone charakteryzowane są przez dwa pojęcia: przybliżenie dolne oraz przybliżenie górne zbioru, których definicje można znaleźć w wielu publikacjach. W systemie bazującym na teorii zbiorów przybliżonych informacje są przechowywane w postaci stabelaryzowanej, co pozwala na łatwe wyodrębnienie atrybutów warunkowych oraz decyzyjnych. Taka forma przechowywania wiedzy umożliwia łatwe wygenerowanie reguł decyzyjnych. Dodatkową zaletą systemu opartego na metodzie zbiorów przybliżonych jest możliwość przetwarzania reguł sprzecznych. Wymaga ona jednak, aby wartości atrybutów miały postać dyskretną, co wiąże się z kwantyzacją parametrów ciągłych.

Filarem projektowanego systemu jest model bazy danych, przedstawiony w formie tabel i powiązań pomiędzy nimi. Główną tabelą są Nieruchomości będące zbiorem obiektów wraz z charakteryzującymi je atrybutami. Można wyróżnić takie atrybuty, jak sygnatura, data wprowadzenia oferty, powierzchnia użytkowa mieszkania (w m²), standard mieszkania, rok budowy, numer piętra, liczba pięter w budynku, otoczenie nieruchomości, dzielnica, w jakiej jest położona, data zawarcia transakcji oraz cena uzyskana za 1 m². Oprócz podstawowej tabeli faktów sprzedaży ujęto atrybuty mieszkania w tabelach

⁸ Z. Matusiewicz, *Systemy informatyczne w nauczaniu zbiorów przybliżonych i rozmytych*, Rzeszów 2006.

słownikowych: Otoczenie, Standard, Dzielnica. Model bazy zawiera też tabelę Otoczenie nieruchomości, łączącą Nieruchomości z Otoczeniem. Otoczenie nieruchomości jest opisane wieloma czynnikami, stąd wykorzystanie tabeli pomocniczej. Atrybutami warunkowymi w opracowanym systemie są następujące cechy nieruchomości: dzielnica, w jakiej leży nieruchomość, numer piętra badanego mieszkania, liczba pięter budynku, powierzchnia użytkowa mieszkania, rok budowy, standard mieszkania. Atrybutem decyzyjnym jest natomiast cena za 1 m² powierzchni użytkowej obiektu. Szczegółowy projekt tabel zawarto na rysunku 1.



Rys. 1. Etapy konstrukcji założeń projektowych systemu

Źródło: opracowanie własne.

Aby system informatyczny spełnił swoje zadanie, musi zostać wyposażony w dane. Z racji tego, że niniejszy artykuł dotyczy modelu systemu eksperckiego ekstrakcji reguł decyzyjnych, techniczny aspekt wprowadzania danych został pominięty. Autorzy tylko sygnalizują ten fakt i przechodzą do kolejnego etapu – dyskretyzacji atrybutów ciągłych. W tym przypadku są to: cena ofertowa oraz powierzchnia użytkowa. Dyskretyzacja polega

na przypisaniu wartości dyskretnej przedziałom ciągłych wartości oryginalnego atrybutu. Można wyróżnić następujące metody dyskretyzacji⁹: dyskretyzacja naiwna, według równej szerokości, według równej częstości, z wykorzystaniem wiedzy eksperta. Dodatkowo wartości atrybutów: numer piętra, liczba pięter oraz rok budowy zostaną podzielone na przedziały tak, aby różnice wartości atrybutów w poszczególnych przedziałach były mało znaczące. Następstwem tych operacji będzie zmniejszony wygenerowany zbiór reguł decyzyjnych, co jest skutkiem zredukowanej liczby różnych wystąpień poszczególnych atrybutów.

Pierwszym atrybutem przeznaczonym do dyskretyzacji jest cena ofertowa za 1 m². Analizując posiadaną bazę danych z informacjami na temat niedawno zrealizowanych transakcji, stwierdzono, że wartości tego parametru znajdują się w przedziale [1245,0; 9000,0], natomiast ich mediana wynosi 2844. Zdecydowano się na podział dziedziny na 7 podzbiorów metodą dyskretyzacji według równej częstości. Otrzymane podzbiory wraz z opisami słownymi przedstawia tabela 2.

Tabela 2

Zdyskretyzowane wartości atrybutu cena ofertowa

$(-\infty;$ 2150,6)	[2150,6; 2435,1)	[2435,1; 2664,1)	[2664,1; 3122,5)	[3122,5; 3895,3)	[3895,3; 4947,3)	[4947,3; $+\infty$)
najniższa	bardzo niska	niska	średnia	wysoka	bardzo wysoka	najwyższa

Źródło: opracowanie własne.

Przeanalizowane wartości powierzchni użytkowej wahają się od 20 do 185 m². W przypadku tego atrybutu zastosowano podział na 5 zbiorów o równej szerokości. Brzegiem dolnym ostatniego podzbioru jest 130, gdyż mieszkania o powierzchni powyżej tej wartości uznawane są za bardzo duże i różnice w ich wymiarach nie są już takie istotne. Wprowadzony podział przedstawia tabela 3.

⁹ Z. Pawlak, *Systemy informacyjne. Podstawy teoretyczne*, Bielsko-Biała 1983.

Tabela 3

Zdyskretyzowane wartości atrybutu powierzchnia użytkowa

$(-\infty; 40,0)$	$[40,0; 70,0)$	$[70,0; 100,0)$	$[100,0; 130,0)$	$[130,0; +\infty)$
bardzo mała	mała	średnia	duża	bardzo duża

Źródło: opracowanie własne.

Następnym atrybutem jest numer piętra, na którym znajduje się mieszkanie. Zaobserwowane przykłady tego atrybutu w dostępnych transakcjach znajdują się w przedziale od 0 (parter) do 12. Wartości te zostały zgrupowane w przedziały z wykorzystaniem wiedzy eksperta (tabela 4).

Tabela 4

Zdyskretyzowane wartości atrybutu numer piętra

$\{0\}$	$(0; 2]$	$(2; 5]$	$(5; 10]$	$(10; +\infty)$
parter	niskie	średnie	wysokie	bardzo wysokie

Źródło: opracowanie własne.

Liczba pięter reprezentowana jest przez podział, jak w przypadku numeru piętra, z tą różnicą, że ten atrybut nie może przyjąć wartości 0. W bardziej zaawansowanym systemie, zakładając, że znany jest numer piętra, na którym znajduje się mieszkanie, oraz ogólna liczba pięter w budynku, można pokusić się o bardziej kompleksową analizę tych atrybutów, przypisując unikalną wartość na przykład ostatniemu numerowi piętra w budynku. Wprowadzone wartości przedstawia tabela 5.

Tabela 5

Zdyskretyzowane wartości atrybutu liczba pięter

$(0; 2]$	$(2; 5]$	$(6; 10]$	$(10; +\infty)$
mała	średnia	duża	bardzo duża

Źródło: opracowanie własne.

Atrybut rok budowy jest specyficznym parametrem, gdyż oprócz wartości liczbowych określających dokładny rok budowy budynku może przyjmować wartości słowne, takie jak „okres przedwojenny”, „lata 70.” itp. W związku z występowaniem tego rodzaju określeń w celu uniknięcia przekłamań szerokość poszczególnych podzbiorów stanowi wielokrotność dekady, a początki i końce tych dekad definiują brzegi przedziału. Wykorzystano tu metodę dyskretyzacji według równej szerokości, dzieląc dziedzinę atrybutu na 5 przedziałów (tabela 6).

Tabela 6

Zdyskretyzowane wartości atrybutu rok budowy

$(-\infty; 1940)$	$[1940; 1960)$	$[1960; 1980)$	$[1980; 2000)$	$[2000; +\infty)$
bardzo wczesny	wczesny	średni	późny	bardzo późny

Źródło: opracowanie własne.

Atrybut tekstowy standard nie wymusza konieczności jego przekształcenia. Przyjmuje on pięć wartości, które dobrze oddają charakter mieszkania, zobrazowane w tabeli 7.

Tabela 7

Wartości atrybutu standard

do remontu	podstawowy	podwyższony	średni	wysoki
------------	------------	-------------	--------	--------

Źródło: opracowanie własne.

Specyficznym atrybutem jest otoczenie. Pojedynczy obiekt może być opisywany jednocześnie przez wiele różnych przykładów tego parametru. Atrybut ten może przyjmować wartości: luźna zabudowa, gęsta zabudowa, osiedle domów jednorodzinnych, osiedle domów wielorodzinnych, park, las, jezioro, centrum handlowe, hale produkcyjne, obiekty przemysłowe. Poszczególnym zbiorom wartości tego atrybutu należy przypisać odpowiednie wartości skali porządkowej, jak zostało to zrealizowane w przypadku wcześniejszych parametrów. Założono, że skala porządkowa otoczenia może przyjąć wartości przedstawione w tabeli 8.

Tabela 8

Wartości atrybutu otoczenie

niekorzystne	umiarkowanie niekorzystne	neutralne	umiarkowanie korzystne	korzystne
--------------	------------------------------	-----------	---------------------------	-----------

Źródło: opracowanie własne.

Po zakończonym procesie dyskretyzacji zbiorów danych podzielono na dwa podzbiory: 433-elementowy zbiór danych uczących oraz 108-elementowy zbiór danych testujących.

Tworzenie zbioru reguł decyzyjnych

Reguła decyzyjna jest wyrażeniem logicznym postaci: *jeżeli* koniunkcja warunków elementarnych, *to* dysjunkcja decyzji elementarnych. Umożliwiają one sformalizowanie prawidłowości zachodzących między danymi i wyrażenie ich w sposób ścisły. Dodatkowo przed rozpoczęciem procesu wnioskowania ze zbioru atrybutów należy wyznaczyć badane pojęcie zwane atrybutem decyzyjnym. Przykładowy fragment tablicy decyzyjnej został przedstawiony w tabeli 9.

Tabela 9

Fragment tablicy decyzyjnej projektowanego systemu

Dzielnica	Pow. użytk.	Numer piętra	Liczba pięter	Standard	Rok budowy	Otoczenie	Cena za 1 m ²
Śródmieście	(*; 40)	(2; 5)	(2; 5)	podwyższony	(*; 1940)	korzystne	[3895,3; 4947,2)
Śródmieście	[40; 70)	(2; 5)	(5; 10]	średni	[1960; 1980)	neutralne	[4947,2; *)
Śródmieście	[70; 100)	(2; 5)	(2; 5)	średni	(*; 1940)	umiarkowanie niekorzystne	[3895,3; 4947,2)
Zachód	(*; 40)	(0; 2]	(0; 2]	średni	[1980; 2000)	neutralne	[4947,2; *)
Zachód	[40; 70)	(0; 2]	(2; 5]	średni	[1980; 2000)	neutralne	[4947,2; *)
Śródmieście	(*; 40)	(2; 5)	(10; *)	podstawowy	[1960; 1980)	neutralne	[3895,3; 4947,2)
Prawobrzeże	[40; 70)	(2; 5)	(2; 5]	średni	[1980; 2000)	neutralne	[4947,2; *)

Dzielnica	Pow. użytk.	Numer piętra	Liczba pięter	Standard	Rok budowy	Otoczenie	Cena za 1 m ²
Śródmieście	[40; 70)	[0; 0]	(2; 5]	średni	[1960; 1980)	neutralne	[3895,3; 4947,2)
Północ	[40; 70)	[0; 0]	(0; 2]	podstawowy	[1960; 1980)	neutralne	[4947,2; *)
Zachód	[40; 70)	[0; 0]	(0; 2]	podwyższony	[2000; *)	neutralne	[4947,2; *)
Śródmieście	[40; 70)	(2; 5]	(2; 5]	średni	[1980; 2000)	neutralne	[3895,3; 4947,2)
Zachód	[70; 100)	(0; 2]	(0; 2]	wysoki	[2000; *)	neutralne	[4947,2; *)

Źródło: opracowanie własne.

W projektowanym systemie atrybutem decyzyjnym jest cena ofertowa za 1 m² powierzchni użytkowej mieszkania. Przedstawione w tabeli 10 dyskretyzowane dane na temat transakcji sprzedażowych nieruchomości w Szczecinie stanowią tabelę decyzyjną, która jest pewną strukturą opisu zbioru związanych ze sobą reguł decyzyjnych. Tablica decyzyjna w bardziej złożonych problemach może przyjmować o wiele większe rozmiary. Istnieje duże prawdopodobieństwo wystąpienia rekordów nadmiarowych, nieistotnych – następstwo tzw. szumu informacyjnego. Problem ten może być rozwiązany za pomocą reduktów. Dla systemu informacyjnego $S = (U, A, V, f)$ oraz $B \subseteq A$, atrybut a nazywamy zbędnym w B , gdy spełniona jest zależność¹⁰:

$$\text{IND}_S(B) = \text{IND}_S(B \setminus \{a\})^2$$

W przeciwnym razie atrybut a nazywamy niezbędnym w B . Zbiór atrybutów B w systemie informacyjnym S jest niezależny, jeżeli dla każdego atrybutu $a \in B$ spełniony jest warunek:

$$\text{IND}_S(B) \neq \text{IND}_S(B \setminus \{a\})^2$$

Zbiór atrybutów B jest niezależny wtedy i tylko wtedy, gdy każdy atrybut należący do B jest niezbędnym w tym zbiorze. W przeciwnym razie zbiór B

¹⁰ P. Nowotarski, D. Chodara, P. Leończyk, *Zbiory przybliżone – wnioskowanie przybliżone*, Tarnobrzeg.

jest zależny. Usunięcie dowolnego atrybutu ze zbioru niezależnego wiąże się z utratą informacji, a więc zmniejszeniem dokładności aproksymacji pojęcia.

Wprowadzone pojęcia pozwalają na sformułowanie definicji samego reduktu. Redukt jest to rodzina zbioru atrybutów wyodrębnionych z pewnego większego zbioru atrybutów, za pomocą których możliwe jest uzyskanie takiej samej dokładności aproksymacji. Istnieją specjalne metody wyznaczenia reduktów: algorytm ewolucyjny, algorytm Johnsona, metoda reduktów dynamicznych, wyczerpujące szacowanie, zmodyfikowany algorytm Johnsona oraz zmodyfikowany algorytm genetyczny¹¹. Do wyznaczenia reduktów dla tablicy decyzyjnej projektowanego systemu do wyceny nieruchomości wykorzystano kolejno wszystkie dostępne metody. W każdym przypadku wynik był ten sam – otrzymany redukt stanowił cały zbiór atrybutów warunkowych znajdujących się w tablicy. Wynika z tego, że wszystkie dostępne parametry charakteryzujące nieruchomość są istotne w procesie wyceny mieszkania.

Mając wyznaczony minimalny redukt, możliwe było wygenerowanie reguł decyzyjnych. Na podstawie 433 obiektów z tablicy decyzyjnej zostało wygenerowanych 318 reguł decyzyjnych. Każda reguła charakteryzowana jest przez tzw. współczynnik dokładności. Dla systemu informacyjnego $S = (U, A, V, f)$ współczynnikiem dokładności pojęcia X nazywamy miarę¹²:

$$\alpha_B(X) = \frac{\text{card}(\underline{B}X)}{\text{card}(\overline{B}X)}$$

Współczynnik dokładności charakteryzuje się następującymi własnościami:

- a) $0 \leq \alpha_B \leq 1$;
- b) jeżeli $\alpha_B(X) = 1$, to pojęcie jest ostre i jego własności mogą być w pełni wyrażone za pomocą zbioru atrybutu B ;
- c) jeżeli $\alpha_B(X) = 0$, to pojęcie jest całkowicie nieostre i jego własności nie mogą być w pełni wyrażone za pomocą zbioru atrybutu B ;
- d) jeżeli $0 < \alpha_B(X) < 1$, to pojęcie jest nieostre i jego własności mogą być częściowo wyrażone za pomocą zbioru atrybutu B ¹³.

¹¹ Z. Pawlak, *Systemy informacyjne. Podstawy teoretyczne...*, dz. cyt.

¹² P. Nowotarski, D. Chodara, P. Leończyk, *Zbiory przybliżone...*, dz. cyt.

¹³ Tamże.

Na podstawie 433 obiektów z tablicy decyzyjnej zostało wygenerowanych 318 reguł decyzyjnych. Przykładowy zestaw wygenerowanych reguł decyzyjnych zawarto w tabeli 10.

Tabela 10

Fragment zbioru reguł decyzyjnych projektowanego systemu

DZ (Śródmieście) AND POW ((*, 40)) AND NR_P ((2, 5]) AND IL_P ((2, 5]) AND ST (podwyższony) AND ROK ((*, 1940)) AND OT (korzystne) => CENA ([3895.3, 4947.2))
DZ (Prawobrzeże) AND POW ([40, 70)) AND NR_P ([0, 0]) AND IL_P ((0, 2]) AND ST (podwyższony) AND ROK ([2000, *]) AND OT (korzystne) => CENA ([4947.2, *])
DZ (Śródmieście) AND POW ([40, 70)) AND NR_P ((2, 5]) AND IL_P ((5, 10]) AND ST (średni) AND ROK ([1960, 1980)) AND OT (neutralne) => CENA ([4947.2, *]) OR CENA ([2435.1, 2664.1))
DZ (Zachód) AND POW ([40, 70)) AND NR_P ((0, 2]) AND IL_P ((0, 2]) AND ST (średni) AND ROK ([2000, *]) AND OT (neutralne) => CENA ([4947.2, *])
DZ (Prawobrzeże) AND POW ([40, 70)) AND NR_P ((2, 5]) AND IL_P ((2, 5]) AND ST (średni) AND ROK ([1980, 2000)) AND OT (korzystne) => CENA ([3895.3, 4947.2))
DZ (Śródmieście) AND POW ([40, 70)) AND NR_P ([0, 0]) AND IL_P ((2, 5]) AND ST (podwyższony) AND ROK ([1980, 2000)) AND OT (neutralne) => CENA ([4947.2, *])
DZ (Śródmieście) AND POW ([70, 100)) AND NR_P ((0, 2]) AND IL_P ((0, 2]) AND ST (wysoki) AND ROK ([2000, *]) AND OT (umiarkowanie korzystne) => CENA ([4947.2, *])
DZ (Śródmieście) AND POW ([40, 70)) AND NR_P ([0, 0]) AND IL_P ((2, 5]) AND ST (średni) AND ROK ([1960, 1980)) AND OT (neutralne) => CENA ([3895.3, 4947.2)) OR CENA ([3122.5, 3895.3))
DZ (Zachód) AND POW ((*, 40)) AND NR_P ((0, 2]) AND IL_P ((0, 2]) AND ST (średni) AND ROK ([1980, 2000)) AND OT (neutralne) => CENA ([4947.2, *])
DZ (Śródmieście) AND POW ([40, 70)) AND NR_P ((2, 5]) AND IL_P ((2, 5]) AND ST (średni) AND ROK ([1940, 1960)) AND OT (neutralne) => CENA ([3895.3, 4947.2))
DZ (Śródmieście) AND POW ((*, 40)) AND NR_P ((2, 5]) AND IL_P ((10, *]) AND ST (podstawowy) AND ROK ([1960, 1980)) AND OT (neutralne) => CENA ([3895.3, 4947.2))
DZ (Śródmieście) AND POW ([40, 70)) AND NR_P ((2, 5]) AND IL_P ((2, 5]) AND ST (do remontu) AND ROK ([1980, 2000)) AND OT (korzystne) => CENA ([4947.2, *])
DZ (Prawobrzeże) AND POW ([40, 70)) AND NR_P ((2, 5]) AND IL_P ((2, 5]) AND ST (średni) AND ROK ([1980, 2000)) AND OT (neutralne) => CENA ([4947.2, *]) OR CENA ([3122.5, 3895.3)) OR CENA ([2150.6, 2435.1)) OR CENA ((*, 2150.6))

DZ (Północ) AND POW ([40, 70]) AND NR_P ([0, 0]) AND IL_P ((0, 2]) AND ST (podstawowy) AND ROK ([1960, 1980]) AND OT (neutralne) => CENA ([4947.2, *])
DZ (Zachód) AND POW ([40, 70]) AND NR_P ([0, 0]) AND IL_P ((0, 2]) AND ST (podwyższony) AND ROK ([2000, *]) AND OT (neutralne) => CENA ([4947.2, *]) OR CENA ([2664.1, 3122.5])
DZ (Śródmieście) AND POW ([40, 70]) AND NR_P ((5, 10]) AND IL_P ((5, 10]) AND ST (średni) AND ROK ([1960, 1980]) AND OT (neutralne) => CENA ([4947.2, *])
DZ (Północ) AND POW ([40, 70]) AND NR_P ((0, 2]) AND IL_P ((2, 5]) AND ST (podwyższony) AND ROK ([1980, 2000]) AND OT (umiarkowanie niekorzystne) => CENA ([4947.2, *])
DZ (Zachód) AND POW ([70, 100]) AND NR_P ((2, 5]) AND IL_P ((2, 5]) AND ST (średni) AND ROK ([1980, 2000]) AND OT (neutralne) => CENA ([4947.2, *]) OR CENA (*, 2150.6)) OR CENA ([3122.5, 3895.3])
DZ (Śródmieście) AND POW ([40, 70]) AND NR_P ((2, 5]) AND IL_P ((2, 5]) AND ST (średni) AND ROK ([1980, 2000]) AND OT (neutralne) => CENA ([3895.3, 4947.2]) OR CENA ([2435.1, 2664.1])
DZ (Zachód) AND POW ([40, 70]) AND NR_P ((5, 10]) AND IL_P ((5, 10]) AND ST (średni) AND ROK ([1960, 1980]) AND OT (neutralne) => CENA ([3895.3, 4947.2]) OR CENA ([2664.1, 3122.5])
DZ (Śródmieście) AND POW ([70, 100]) AND NR_P ((2, 5]) AND IL_P ((2, 5]) AND ST (średni) AND ROK ([1980, 2000]) AND OT (umiarkowanie korzystne) => CENA ([4947.2, *])
DZ (Śródmieście) AND POW ([40, 70]) AND NR_P ((2, 5]) AND IL_P ((5, 10]) AND ST (podwyższony) AND ROK ([1980, 2000]) AND OT (neutralne) => CENA ([4947.2, *])
DZ (Prawobrzeże) AND POW ([40, 70]) AND NR_P ([0, 0]) AND IL_P ((2, 5]) AND ST (wysoki) AND ROK ([2000, *]) AND OT (korzystne) => CENA ([4947.2, *])
DZ (Północ) AND POW ([70, 100]) AND NR_P ([0, 0]) AND IL_P ((0, 2]) AND ST (podstawowy) AND ROK ([2000, *]) AND OT (korzystne) => CENA ([4947.2, *])

Źródło: opracowanie własne.

Na podstawie współczynników dokładności możliwe jest usunięcie reguł najmniej wiarygodnych. W czasie projektowania systemu zastosowano wiele metod, jednak żadna z nich nie zredukowała liczby wygenerowanych wcześniej reguł.

Podsumowanie

Głównym tematem artykułu jest model systemu ekspertowego ekstrakcji reguł decyzyjnych dla potrzeb wyceny nieruchomości. Przedstawiono tradycyjne metody wyceny wraz z ich podstawowymi założeniami oraz niezbędnymi warunkami wyznaczającymi możliwości ich wykorzystania. Wykazano, iż dopuszczalne sytuacje wykorzystania poszczególnych sposobów wyceny nieruchomości są mocno ograniczone ich wymiarem praktycznym oraz regulacjami prawnymi. Przedstawiono model systemu ekspertowego wykorzystującego teorię zbiorów przybliżonych w procesie generowania reguł decyzyjnych dla potrzeb wyceny nieruchomości. Projekt zrealizowano, wzbogacając poszczególne etapy procesu modelowania o formalne definicje oraz zależności. Przedstawiono kolejne etapy budowy systemu: model bazy danych, przetworzenie zebranych danych, aż do pozyskania zbioru reguł decyzyjnych zasilających bazę wiedzy systemu. Przedstawione wyniki symulacji wskazują, że celowe jest wykorzystanie systemów ekspertowych do wyceny nieruchomości. Zalety inteligentnych narzędzi wspomagających podejmowanie decyzji są nieocenione, jednak narzędzia te nadal wymagają udoskonaleń w wielu aspektach.

Dalszym kierunkiem badań nad zwiększeniem efektywności klasyfikowania prezentowanego systemu powinno być skonstruowanie algorytmu umożliwiającego odkrywanie wiedzy dotyczącej brakujących wartości atrybutów warunkowych przy jednoczesnym zwiększeniu zbioru parametrów charakteryzujących nieruchomość uwzględnianych w procesie wnioskowania. Obszerniejsza i bardziej szczegółowa wiedza implikuje zwiększoną jakość wnioskowania przy jednoczesnym zachowaniu zdolności do jego generalizowania.

Literatura

AFR Insurance Services – Automated Valuation Models, <<http://insurance.afrservices.com/Elite/contentmgmt/ContentItemView.aspx?ItemID=3dc28627-7cb3-4a1d-bdc8-60c5e82f021f>>.

Automated Valuation Models, <<http://www.appraisaltoday.com/avms.htm>>.

Downie M.L., Robson G., *Automated Valuation Models: an international perspective*, Londyn 2007.

Internetowy Podręcznik Statystyki, <<http://www.dataminer.pl/textbook/stnaiveb.html>>.

- Isaac D., Crosby N., *Metody wyceny nieruchomości*, Hongkong 1991.
- Kucharska-Stasiak E., *Nieruchomość a rynek*, Warszawa 1999.
- Matusiewicz Z., *Systemy informatyczne w nauczaniu zbiorów przybliżonych i rozmytych*, Rzeszów 2006.
- Mulawka J., *Systemy ekspertowe*, Warszawa, 1997.
- Nowotarski P., Chodara D., Leończyk P., *Zbiory przybliżone – wnioskowanie przybliżone*, Tarnobrzeg 2004.
- Ohm A., *ROSETTA Technical Reference Manual*, Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim 2001.
- Pawlak Z., *Systemy informacyjne. Podstawy teoretyczne*, Wydawnictwo Naukowo-Techniczne, Warszawa 1983.
- Rough Set Techniques in Knowledge Discovery and Data Mining*, The Ninth Pacific-Asia Conference on Knowledge Discovery and Data Mining, Hanoi, Vietnam 2005.
- Śliwiński A., *Zarządzanie nieruchomościami: podstawy wiedzy i praktyki gospodarowania nieruchomościami*, wyd. I, Warszawa 2000.
- Ustawa o gospodarce nieruchomościami, <<http://e-prawnik.pl/kodeksy/ustawy/ustawa-o-gospodarce-nieruchomosciami-art-149-242>> [data dostępu: 20.11.2009].
- ValuePoint®4 – *First American CoreLogic, Inc.*, <<http://www.facorelogic.com/literature/valuepoint4.jsp>> [data dostępu: 24.11.2009].

AN EXPERT MODEL OF DECISION RULES EXTRACTION FOR THE COST ESTIMATE OF ESTATES

Summary

The main aim of this article is the proposal of expert model of decision rules extraction for cost estimate of estates. The traditional methods of cost estimate and their postulates and necessary conditions of their exploitations were proposed. This article presents the model of expert system using the fuzzy set theory in generation process of decision rules for cost estimate of estates. In this project, phases of process modeling were enriched on formal definitions and connections. The results of this simulation prove that the using of expert systems for cost estimate of estates is a relevant method.

Translated by Agnieszka Konys