

Krzysztof ZIMA¹

SZACOWANIE WSTĘPNYCH KOSZTÓW BUDOWY BOISK SPORTOWYCH Z WYKORZYSTANIEM WNIOSKOWANIA Z PRZYPADKÓW

W artykule zostanie pokazane zastosowanie metody CBR (Case Based Reasoning) na wczesnym etapie budowlanego procesu inwestycyjnego w celu szacowania kosztów boisk sportowych. Analiza ma na celu zbadanie możliwości wspomagania kosztorysowania robót budowlanych z wykorzystaniem wnioskowania z przypadków. Zaprezentowana zostanie baza danych zawierająca 84 przypadki boisk sportowych dla których przeprowadzono postępowania przetargowe. Baza danych zawiera także ceny wybranych ofert w toku postępowania. Następnie pokazano proces wnioskowania z przypadków w celu wyboru najbardziej zbliżonych przypadków do analizowanego problemu. Obliczone zostały lokalne podobieństwa przypadków nowych i starych zawartych w bazie danych, a następnie podobieństwo globalne. Wybrane zostały przypadki, dla których podobieństwo globalne było większe niż 95%. Wybrane przypadki zostały następnie adaptowane ze względu na różnice w czasie kalkulowanych ofert starych przypadków, a czasem oszacowania dla nowego przypadku oraz różnic w lokalizacji inwestycji. W tym celu dokonano korekty o dwa współczynniki korekcyjne: współczynnik waloryzacyjny i współczynnik regionalny. Obliczenia przeprowadzono dla czterech modeli różniących się przyjętymi wagami zmiennych objaśniających przypadki oraz ilością zmiennych dla których obliczane były podobieństwa lokalne. Na koniec wyliczono średni absolutny szacowany błąd MAEE dla analizowanych modeli i dokonano analizy błędów w odniesieniu do publikowanych wskaźników polskich oraz zagranicznych w celu oceny błędu dla analizowanych modeli.

Słowa kluczowe: Case Based Reasoning, kosztorysowanie, roboty budowlane

1. Wprowadzenie

Metoda wnioskowania z przypadków (Case Based Reasoning – CBR) jest często wykorzystywana w rozwiązywaniu problemów decyzyjnych. Wnioskowanie na podstawie przypadków wykorzystuje wiedzę specyficzną zawartą w doświadczonych w przeszłości sytuacjach, zwanych przypadkami [2].

¹ Krzysztof Zima, Politechnika Krakowska, Wydział Inżynierii Lądowej, Instytut Zarządzania w Budownictwie i Transporcie, ul. Warszawska 24, 31-155 Kraków, kzima@izwbit.pk.edu.pl

Wnioskowanie na podstawie znanych przypadków to metodologia wykorzystująca proste spostrzeżenia że „podobne problemy już kiedyś rozwiązano” oraz „podobne przypadki mają podobne rozwiązania” [10]. CBR imituje, w bardziej wierny sposób, ludzką inteligencję i proces nauki ponieważ aplikowany system modyfikuje swoje zachowanie na podstawie zgromadzonego doświadczenia [9]. Pierwszym krokiem w metodzie CBR jest analiza rozważanego problemu. Kolejnym krokiem jest znalezienie przypadków, które są najbardziej podobne do analizowanego przypadku. Zdarza się że żaden stary przypadek nie pasuje do nowego. Należy wtedy wprowadzić modyfikacje w starych rozwiązaniach. Następnym krokiem jest dokonanie oceny uzyskanych rozwiązań. Proponowane wybrane rozwiązanie jest testowane, a następnie są dokonywane niezbędne poprawki. Metodologia wnioskowania z przypadków może być szeroko wykorzystywana w różnych dziedzinach gospodarki. Może być również stosowana w budownictwie: w procesie inżynierii produkcji, oszacowaniach kosztowych oraz tworzeniu procedur zarządzania jakością [7]. Metoda CBR może być wykorzystywana w budownictwie między innymi do celów zarządzania kosztami przedsięwzięcia lub przygotowania wiarygodnych oszacowań kosztowych dla inwestora, odzwierciedlających ceny rynkowe [11].

W oszacowaniach kosztowych dokonywanych we wstępnej fazie inwestycji stosunkowo często wykorzystywana jest metoda wnioskowania z przypadków. Przykładowo w [5] autorzy opisali model hybrydowy wykorzystujący wnioskowanie z przypadków wspomagane algorytmami genetycznymi. Zaprezentowano model szacowania kosztów wodnych obiektów budowlanych oparty się na bazie danych złożonej ze 189 przypadków, a po zbadaniu korelacji z kosztami inwestycji wykorzystali 6 zmiennych opisujących inwestycje. Z kolei An, Kim i Kang [1] wsparli proces wnioskowania z przypadków analitycznym procesem hierarchicznym AHP. Kim i Shim [6] opisali system hybrydowy oparty na CBR i algorytmach genetycznych. Algorytmy genetyczne zostały wykorzystane do optymalizacji określenia wag kryteriów. Wykorzystana w obliczeniach baza danych zawierała 590 budynków zrealizowanych w Korei Południowej w latach 2000-2010. Podobny sposób zaprezentowali dla przedsięwzięć budowlanych Ji, Park i Lee [3]. Z kolei Marzouk i Ahmed [8] wykorzystali metodę CBR do szacowania kosztów stacji pomp. Wykorzystali w tym celu bazę danych złożoną z 44 inwestycji opierając się na 14 kryteriach.

2. Modele szacowania wstępnych kosztów z wykorzystaniem wnioskowania z przypadków

2.1. Baza danych boisk sportowych

Utworzona przez autora baza danych zawiera 84 przypadki inwestycji polegających na budowie boisk sportowych. Zebrane informacje pochodzą z zamówień publicznych opisanych kodem CPV - 45212200-8 „Roboty budowlane

w zakresie budowy obiektów sportowych”. Dane pochodzą z wybranych w toku postępowania przetargowego ofert z 2014 r. i dotyczą robót budowlano-inwestycyjnych na wykonanie boisk sportowych. Baza danych boisk sportowych BD_{BS} zawiera przypadki zawierające 3 grupy informacji R_i , OS_i , O_i :

$$BD_{BS} = \{\text{Przypadek}_1[(R_1), (OS_1), (O_1)], \dots, \text{Przypadek}_{84}[(R_{84}), (OS_{84}), (O_{84})]\} \quad (1)$$

gdzie:

Przypadek_i [R_i , OS_i , O_i], – i-ty „stary” przypadek,

R_i – rozwiązanie dla i-tego przypadku,

OS_i – opis sytuacji dla i-tego przypadku,

O_i – zmienne objaśniające dla i-tego przypadku.

Rozwiązaniem jest cena jednostkowa m^2 powierzchni boiska sportowego wyrażona w $zł/m^2$. Opis sytuacji zawiera niezbędne informacje opisujące sytuację takie jak: funkcja boiska, lokalizacja oraz data oferty. Zmienne objaśniające to dane ilościowe zawierające informacje o powierzchniach w $[m^2]$ boisk, dojeżdż i dojazdów, zieleni oraz powierzchni piłkochwyków, czy długości w $[m]$ ogrodzeń oraz dane jakościowe opisujące typ nawierzchni sportowej, rodzaj materiału na dojeżdża i dojazdy, typ ogrodzenia oraz rodzaj wyposażenia sportowego.

Fragment bazy danych (kolejnych rekordów) został przedstawiony poniżej: Przypadek_n [(cena jednostkowa); (funkcja boiska; lokalizacja; data oferty); (powierzchnia boiska; powierzchnia dojeżdż i dojazdów; powierzchnia zieleni; powierzchnia piłkochwyków, długość ogrodzenia; typ nawierzchni sportowej; materiał na dojeżdża i dojazdy; typ ogrodzenia; rodzaj wyposażenia w notacji 0/1=NIE/TAK – piłka ręczna-siatkówka-koszykówka-piłka nożna-tenis ziemny)],

Przypadek₁ [(321,07 $zł/m^2$); (wielofunkcyjne, Bytom, 24.02.2014), 1196 m^2 , 732 m^2 , 640 m^2 , 240 m^2 , 144 m; poliuretanowa, kostka brukowa, panelowe, 11101],

Przypadek₂ [(280,54 $zł/m^2$); (wielofunkcyjne, Dąbrówka, 28.02.2014, 392 m^2 , 0 m^2 , 1867 m^2 , 0 m^2 , 86 m; kostka betonowa bezfazowa, brak, panelowe, 11101],

Przypadek₃ [(8,94 $zł/m^2$); piłka nożna, Biskupie, 04.03.2014, 2970 m^2 , 0 m^2 , 0 m^2 , 0 m; trawa naturalna, brak, brak, 00010].

2.2. Wnioskowanie z przypadków – przykład obliczeniowy

Analizie poddano przypadki, które do celów badawczych potraktowano jako nowe V_N – 4 przypadki z roku 2015 i 4 przypadki z roku 2016. Przypadki wraz z opisem sytuacji i zmiennymi objaśniającymi pokazano w tab. 1.

W pierwszym kroku dokonano analizy podobieństwa na poziomie lokalnym dla poszczególnych zmiennych objaśniających.

Tabela 1. Fragment zbioru przypadków testujących wraz z opisem sytuacji i objaśnieniami

Table 1. Fragment of test cases set, with a description of the situation and explanations

Objaśnienia	J.m.	V _{N1}	...	V _{N8}
Cena	[zł]	419 999,50	...	666 778,03
Cena jednostkowa	[zł/m ² boiska]	336,00	...	688,41
Data	-	02.06.2015	...	25.02.2016
Lokalizacja	-	Stopnica	...	Iłża
Powierzchnia boisk	[m ²]	1250	...	968,58
Powierzchnia dojeżdż	[m ²]	105	...	210
Powierzchnia zieleni	[m ²]	36.5	...	150
Długość ogrodzenia	[m]	168	...	258
Powierzchnia piłkochwyków	[m ²]	0	...	0
Przeznaczenie	-	wielofunkcyjne	...	wielofunkcyjne
Typ nawierzchni	-	poliuretanowa	...	poliuretanowa
Materiał na dojścia	-	kostka brukowa	...	kostka brukowa
Rodzaj wyposażenia	0/1	10101	...	11100
Typ ogrodzenia	-	z siatki	...	panelowe

W celu pomiaru lokalnych podobieństw wykorzystano poniższe formuły:

- dla zmiennych objaśniających ilościowych

$$\text{sim}(V_N, V_S) = 1 - \frac{|V_N - V_S|}{V_{\max} - V_{\min}} \quad (2)$$

- dla zmiennych objaśniających jakościowych

$$\text{sim}(V_N, V_S) = 1 - \frac{|n(V_N) - n(V_S)|}{M - 1} \quad (3)$$

gdzie:

$n(V_N)$, $n(V_S)$ – oceny wartości zmiennych objaśniających w przyjętej skali,
 V_{\max} , V_{\min} – wartości minimalne i maksymalne w bazie danych dla danej zmiennej objaśniającej.

W przypadku doboru skali oceny dla zmiennych jakościowych problemem jest dobór liczby poziomów skali, prawidłowość oceny rozwiązań zapisanych za pomocą skali liczbowej, a co za tym idzie ocena podobieństwa rozwiązań.

Miarę $\text{SIM}(V_N, V_S)$ nazywamy globalnym podobieństwem przypadku nowego V_N do przypadku starego V_S . Miary lokalne poszczególnych podobieństw lokalnych $\text{sim}(V_N, V_S)$ są korygowane o wagi zmiennych objaśniających:

$$\text{SIM}(V_N, V_S) = \sum_{i=1}^n \omega_i (\text{sim}_i(V_{Ni}, V_{Si})) \quad (4)$$

gdzie:

ω_i – waga i-tej zmiennej objaśniającej.

Przykład obliczania podobieństw lokalnych $\text{sim}(V_N, V_S)$ oraz podobieństwa globalnego $\text{SIM}(V_N, V_S)$ został przedstawiony w tabeli 2.

Tabela 2. Przykład obliczania podobieństw lokalnych $\text{sim}(V_{N1}, V_{S1})$ oraz podobieństwa globalnego $\text{SIM}(V_{N1}, V_{S1})$

Table 2. An example of calculating local similarities $\text{sim}(V_{N1}, V_{S1})$ and global similarity $\text{SIM}(V_{N1}, V_{S1})$

Zmienne objaśniające	Nowy przypadek 1	Stary przypadek 1	Podobieństwo sim
1. Powierzchnia boisk	968,58 [m ²]	1196 [m ²]	0,98
2. Powierzchnia dojeżdż	210 [m ²]	732 [m ²]	0,91
3. Powierzchnia zieleni	150 [m ²]	640 [m ²]	0,90
4. Długość ogrodzenia	258 [m]	144 [m]	0,88
5. Powierzchnia piłkochwytów	0 [m ²]	240 [m ²]	0,89
6. Przeznaczenie	wielofunkcyjne	wielofunkcyjne	1
7. Typ nawierzchni	poliuretanowa	poliuretanowa	1
8. Materiał na dojeżdż	kostka brukowa	kostka brukowa	1
9. Rodzaj wyposażenia	11100	11101	0,80
10. Typ ogrodzenia	panelowe	panelowe	1
Podobieństwo globalne SIM			0,940

W wyniku analizy wybrane zostały stare przypadki V_S^* , których podobieństwo globalne $\text{SIM}(V_{N1}, V_{S1})$ wynosiło minimum 95%. Każdy z wybranych przypadków został poddany adaptacji poprzez korektę ceny jednostkowej znalezionych przypadków z bazy danych BD_{BS} o dwa współczynniki:

1. Współczynnik regionalny – odzwierciedlający różnice cenowe pomiędzy lokalizacjami inwestycji przypadku nowego V_N oraz wybranego podobnego przypadku starego V_S^* ,
2. Współczynnik waloryzacyjny - odzwierciedlający różnice cenowe spowodowane różnicą w czasie pomiędzy starym V_S^* , a nowym przypadkiem V_N .

Jeżeli w wyniku analizy otrzymano dwa lub więcej przypadków spełniających warunek $\text{SIM}(V_N, V_S) \geq 95\%$ to cena jednostkowa C_N nowego przypadku V_N jest średnią arytmetyczną wybranych przypadków. Przypadek dla którego cena jednostkowa $C_i > 25\%$ od średniej ceny jednostkowej C_N jest odrzucany z wyjątkiem sytuacji gdy mamy do czynienia jedynie z dwoma wybranymi przypadkami. Przy braku rozwiązań spełniających nierówność $\text{SIM}(V_N, V_S) \geq 95\%$ wybierany jest przypadek o najwyższej wartości podobieństwa.

2.3. Analiza dokładności szacowania wstępnych kosztów budowy boisk

W celu analizy utworzono cztery modele: CBR 1, CBR 2, CBR 3 oraz CBR 4. Model CBR 1 zakłada wnioskowanie z przypadków oparte na 10 zmiennych objaśniających pokazanych w tabeli 2 przy założeniu jednakowych wag dla każdego kryterium. Modele CBR 2 i CBR 3 zakładają zróżnicowanie wag poszczególnych zmiennych objaśniających. Wagi zostały dobrane subiektywnie. Ostatni model CBR 4 jest oparty na 6 zmiennych (zmienne objaśniające nr 1, 2, 4, 7, 8 oraz 10). Zmienne te zostały wybrane w toku analizy AHP oceniającej ich ważność w zakresie wpływu na koszt budowy boisk sportowych.

W tabeli 3 zostały zaprezentowane wyniki analizy oszacowania ceny jednostkowej z wykorzystaniem metody CBR. Oceny dokonano na podstawie obliczonego średniego absolutnego szacowanego błędu MAEE. Oszacowano również błędy dla poszczególnych, pojedynczych, analizowanych nowych przypadków. Przypadek 6 był specyficzny i nie było takich przypadków w bazie stąd duży błąd obliczeń. Problemem powodującym wysokie błędy procentowe jest fakt przyjętego poziomu szczegółowości oceny wariantów (wybrane kryteria) i co za tym idzie możliwość występowania w porównywanych boiskach elementów, których nie ma w innych porównywanych boiskach.

Błąd bez Przypadku 6 dla modeli wyniósł kolejno: CBR 1=15,87%, CBR 2=13,24%, CBR 3=12,04%, CBR 4=50,98%. Błąd sumaryczny MAEE powinien wynosić 0 lub być do zera jak najbardziej zbliżony. Jednakże błąd oszacowań kosztowych w fazie wyboru wykonawcy jest w zamówieniach publicznych na poziomie 20-30%. Błąd oszacowania niższy niż 20% uznano za pożądany. Błąd jest więc dopuszczalny (z wyjątkiem modelu CBR 4). Spełnione są też wymogi AACE (American Association of Cost Engineers) dla modeli CBR 2 i CBR 3 – przedział błędu od -10% do + 15% oraz Instytutu Przemysłu Budowlanego (Construction Industry Institute) wynoszące -30% do + 50% (za: [6]). Warto zauważyć, że spełnione są także założenia przyjęte w publikacji [4] dla oszacowań w fazie przedprojektowej takich jak: preliminarz kosztów $\pm 25\%$ -35% oraz budżet $\pm 15\%$ -25%.

Tabela 3. Średni absolutny szacowany błąd MAEE dla analizowanych czterech modeli CBR

Table 3. Mean absolute estimate error MAEE for analyzed four CBR models

Przypadki	Model CBR 1	Model CBR 2	Model CBR 3	Model CBR 4
Przypadek 1	(-) 29,50%	(-) 24,29%	(-) 24,29%	(-) 22,68%
Przypadek 2	(-) 17,26%	(-) 24,95%	(-) 17,26%	(+) 122,64%
Przypadek 3	(+) 2,17%	(+) 2,17%	(+) 2,17%	(+) 152,63%
Przypadek 4	(+) 20,00%	(+) 16,83%	(+) 16,83%	(+) 15,38%
Przypadek 5	(+) 5,81%	(+) 5,81%	(+) 5,81%	(+) 5,81%
Przypadek 6	(+) 422,33%	(+) 294,84%	(+) 294,84%	(+) 620,49%
Przypadek 7	(-) 13,17%	(-) 13,17%	(-) 13,17%	(-) 63,34%
Przypadek 8	(+) 39,05%	(+) 18,72%	(+) 16,77%	(+) 25,40%
Błąd sumaryczny	68,66%	50,10%	48,89%	128,55%
Błąd sumaryczny bez Przypadku 6	15,87%	13,24%	12,04%	50,98%

3. Wnioski

Najlepsze rozwiązania generuje model CBR 3, nieco gorsze modele CRB 1 i CRB 2. Zauważyć można niską wrażliwość modelu na zmiany wag zmiennych. Wyniki nie akceptowalne wygenerował jedynie model CBR 4. Model wykorzystujący jedynie 6 zmiennych sprawdza się głównie w przypadku dużej liczby wyników przekraczających 95% podobieństwa. W przypadku małej liczby przypadków (1-2 przypadki) spełniających warunek 95% podobieństwa model może generować bardzo duże błędy (Przypadki 2,3,7).

Zaletami metody CBR są m.in. czas rozwiązania problemu, brak konieczności dogłębnego zrozumienia problemu oraz możliwość ciągłego uczenia się (przyrost bazy danych i adaptacja przypadków). Metoda CBR ma jednak także wady takie jak: trudność znalezienia odpowiednich przypadków w sytuacji kiedy dysponujemy małą bazą danych (vide Przypadek 6) lub mamy do czynienia z problemem opisywanym przez dużą liczbę zmiennych. Mała baza danych powoduje także konieczność adaptacji znalezionych rozwiązań z uwagi na trudność znalezienia przypadków o 100% podobieństwie. Problemem jest także wybór odpowiednich zmiennych opisujących problem i oszacowanie wag poszczególnych zmiennych. W przypadku szacowania kosztów dochodzi jeszcze czynnik czasu powodujący konieczność stosowania współczynników waloryzacyjnych. Niemniej jednak wyniki oszacowania kosztów we wstępnej fazie inwestycji są satysfakcjonujące dla modeli CBR 1, CBR 2 i CBR 3.

Literatura

- [1] An S-H., Kim G-H., Kang K-I.: A case-based reasoning cost estimating model using experience by analytic hierarchy process, *Building and Environment*, vol. 42, no. 7, 2007, pp. 2573–2579.
- [2] Dohn K., Gumiński A., Matusek M., Zoleński W.: Model wspomaganie zarządzania w zakresie zarządzania wiedzą w polskich przedsiębiorstwach budowy maszyn, *Di-fin*, Warszawa 2013.
- [3] Ji S-H., Park M., and Lee H-S.: Cost estimation model for building projects using case-based reasoning, *Canadian Journal of Civil Engineering*, vol. 38, no. 5, 2011, pp. 570–581.
- [4] Kasprowicz T.: Inżynieria przedsięwzięć budowlanych w: Kapliński O. (red.), *Metody i modele badań w inżynierii przedsięwzięć budowlanych*, KILiW PAN, Warszawa 2007.
- [5] Kim M., Lee S., Woo S., and Shin D.H.: Approximate Cost Estimating Model for River Facility Construction Based on Case-Based Reasoning with Genetic Algorithms, *KSCE Journal of Civil Engineering*, vol. 16, no. 3, 2012, pp. 283-292.
- [6] Kim S., Shim J.H.: Combining case-based reasoning with genetic algorithm optimization for preliminary cost estimation in construction industry, *Canadian Journal of Civil Engineering*, vol. 41, no. 1, 2014, pp. 65-73.
- [7] Kończak A., Paślawski J.: Hybrid approach in learning from examples in construction process design, *Technical Transactions, Civil Engineering*, 2-B/2014, pp. 21-30.

- [8] Marzouk M., Ahmed R.: A case-based reasoning approach for estimating the costs of pump station projects, *Journal of Advanced Research*, vol. 2, no. 4, 2011, pp. 289–295.
- [9] Radziejowska A., Zima K.: The concept of a knowledge base to aid in cost estimating of sports facilities, *International Journal of Contemporary Management*, vol. 14, no. 3, 2015, pp. 99–113.
- [10] Traczyk W.: *Inżynieria wiedzy*, Exit, Warszawa 2010.
- [11] Zima K.: The Case-Based Reasoning model of cost estimation at the preliminary stage of a construction project, *Procedia Engineering*, vol. 122, 2015, pp. 57–64.

EARLY CONSTRUCTION SPORTS FIELDS COST ESTIMATION USING CASE BASED REASONING

Summary

In the article was shown the method Case Based Reasoning - CBR at an early stage of construction investment process in order to estimate the cost of sports fields. The analysis aims to explore the possibility of supporting cost estimation of construction works using CBR. It was presented a database containing 84 cases of sports fields for which tenders were conducted. The database also contains prices of selected offers in the course of the proceedings. Then, has been illustrated application process of the cases in order to select the most similar case to the analyzed problem. Have been calculated local similarities of cases, new and old cases, in the database, and then the global similarity. Cases were chosen for which the global similarity was greater than 95%. Selected cases have subsequently been adapted due to the differences in the calculated time of old offers and a time of calculation for the new case and differences in the location of investments. For this purpose, adjustments were made by two correction factors: coefficient of valorisation and regional coefficient. The calculations were performed for four models differing in the adopted weighting variables for cases and the number of variables for which were calculated local similarities. At the end has been calculated average absolute estimate error MAEE for the analyzed models and errors were analyzed in relation to the published Polish and foreign indicators to assess error for the analyzed models.

Keywords: Case Based Reasoning, cost estimation, construction works

Przesłano do redakcji: 07.06.2016 r.

Przyjęto do druku: 30.06.2016 r.

DOI: 10.7862/rb.2016.26