

**МОДЕЛ ЗА АНТИЦИПАЦИЈУ ЦЕНА НЕКРЕТНИНА  
ЗАСНОВАН НА КОНЦЕПТУ ЗОС И ЛА  
MODEL FOR ANTICIPATION OF REAL ESTATE PRICES  
BASED ON CBR AND LA**

Милан Мркаљ

**REZIME:** Предмет овог рада је развој модела за антиципацију цена некретнина заснованог на концепту закључивања на основу случаја (ЗОС) уз употребу интерполативне Булове алгебре (ИБА) и логичке агрегације (ЛА) као оператора агрегације мере сличности, који за задате улазне вредности атрибута некретнине проналази најсличнију из базе знања - прототипова и према њој процењује цену, као и упоређивање резултата са резултатима базираним на коришћењу класичног ЗОС и агрегацији преко Еуклидске норме. Приказани су општи концепти логичке агрегације и закључивања на основу случаја као и модел настао прилагођавањем ових концепата конкретном проблему.

**KLJUČNE REČI:** моделовање цена некретнина, закључивање на основу случаја (ЗОС), интерполативна Булова алгебра (ИБА), логичка агрегација (ЛА)

**ABSTRACT:** Main goal of this work is the development of model for anticipation of real estate prices based on case based reasoning (CBR) with implementation of logical aggregation (LA) operator derived from Interpolative Boolean Algebra (IBA) and comparison of the results with results based on usage of Euclidean norm. The model input are the values of quality attributes of a real estate, the most similar one is being found from knowledge base, and according to it, the price is being estimated from the input. General concepts of CBR and LA are shown and the model efluents as an adjustment of described general concepts applied onto the problem.

**KEY WORDS:** modeling real-estate prices, Case-Based Reasoning (CBR), Interpolative Boolean Algebra (IBA), Logical Aggregation (LA)

#### УВОД

Поуздано предвиђање цена некретнина је важно за потенцијалне власнике кућа, грађевинске извођаче, инвеститоре, процењиваче, опорезиваче и друге учеснике тржишта некретнина, као што су повериоци и осигуравајућа друштва. Традиционалној методи предвиђања цена некретнина недостаје општеприхваћена стандардизација и сертификација. Услед тога, доступност доброг модела за предвиђање цена некретнина попуњава празнину у информацијама и стога повећава ефикасност тржишта некретнина. Иначе општеприхваћено схватање је да је куповина некретнине један од најкориснијих и најпрофитабилнијих облика инвестирања [5].

Тржиште некретнина у Београду почело је да се формира почетком деведесетих година, када је почео масован откуп друштвених станова. Све до тада промет се одвијао спорадично, а процене су да је станова у власништву било само око 5 одсто. Цене су биле формиране по нетржишним условима. Цена квадратног метра се у моменту уговарања кретала око 2,000 DEM, али је због инфлације у тренутку усељења била много мања. После откупа, због рата, наглог исељавања и пратећих збивања, почела је масовна распродаја некретнина, тако да је то био тренутак најниже цене. Тада је квадратни метар у Новом Београду био испод 1,000 DEM. Од 1994. па надаље, цене непрекидно расту (не рачунајући период рата 1999. ).

У поређењу са другим тржиштима на којима односи понуде и потражње (углавном) регулишу цене, количине промета, на тржишту некретнина код нас неки други

механизми регулишу цене. Кључне речи су очекивања и предвиђања, али не адаптивна (заснована на искуствима из прошлости), нити рационална (темељена на свим расположивим научним информацијама из прошлости и о будућности), како то теорија очекивања дефинише, него готово ирационална која вековима познате економске међудоносе доводе у питање. Оно што донекле ублажава осећај немоћи и несналажења јесте чињеница да је и у другим земљама слично, нарочито у Европи. Цене станова и кућа необично су снажно склоне спекулацијама, продаја и куповина надасве су отежани.

Недоступност потпуних информација које би довеле до квалитетних скупова података у бази знања има за последицу да модели у овом раду имају ограничења по питању прецизности (неки атрибути који имају значајан фактор утицаја на цену некретнине не могу бити уграђени у базу знања). Такође, цене некретнина из базе знања нису сасвим тржишно нивелисане, јер се не ради о ценама извршених купопродаја (базе података са овим ценама нису доступне јавности), већ о ценама актуелне понуде.

Упркос ограничењима, експериментисањем на доступном скупу података из базе знања и развојем модела (различит избор тежинских коефицијената, оператора агрегације и типова генерализованих производа), дошло се до резултата у виду два модела са солидним могућностима процене/предвиђања цена некретнина.

Једно од решења за моделирање цена некретнина је хедонички приступ у коме тржишно препознате субјективне преференције (од стране купаца) регулишу цене [5].

У последње две деценије било је бројних емпиријских студија које су анализирале вредности стамбених објеката. Свака наредна студија је генерално повећавала моћ предвиђања модела наглашавајући микролокацију, квалитет одржавања, макролокацију (географску локацију) и окружење. Неке од новијих студија су додатно фокусиране на трошкове трансакција, као и будуће трошкове произашле по основу поседовања некретнине. Ограничен број студија користи неуронске мреже. У једној студији [5] урађен је упоредни приказ хедоничког типа моделирања комбинован са употребом вештачке неуронске мреже, како би се утврдио најадекватнији метод за конкретан проблем предвиђања цена некретнина у једној области.

Хедонички приступ је адекватан за прилагођавање различитости атрибута при моделирању цена, док његова употреба није реалистична уколико се нека географска област узима као једна територијална “јединица”. Потребно је у модел унети и атрибут који описује макролокацију, јер се ради о високо корелисаном атрибуту са ценом [5].

У овом раду развијен је модел базиран на концептима ЗОС, уз агрегацију мере сличности преко оператора агрегације изведеног из концепта логичке агрегације засноване на ИБА, што представља до сада јединствени начин приступа моделирању цена некретнина и омогућује висок степен адаптације модела природи конкретне проблеме. У модел су укључени атрибуту чије су квантитативне мере посредно или непосредно доступне, а од значаја су за моделирање цене. Модел може бити добра основа за развој web апликације која би имплементирала структуру модела и процесирала квантитативне нормализоване мере атрибута на вредносном нивоу и генерисала апроксимирани цене и тако омогућила широком кругу интересних и заинтересованих страна да повећају ефикасност тржишта некретнина.

Резултати овог модела су упоређени са резултатима класичног приступа ЗОС уз агрегацију мере сличности преко Еуклидске норме. Нови (хибридни) модел се на ограниченом узорку показао бољим од модела заснованог на класичном ЗОС приступу што је последица веће адекватности и прилагођености хибридног модела конкретном проблему.

### 1. ОПИС ОПШТЕГ КОНЦЕПТА МОДЕЛА ЛОГИЧКЕ АГРЕГАЦИЈЕ

Анализиран проблем је агрегација квантитативних мера више особина (парцијалних мера сличности) у једну јединствену меру ради успостављања линеарног једнозначног поретка. У пракси агрегација је најчешће тривијална и као последица тога неадекватна. У овом раду агрегација је третирана као логичка или псеудо-логичка операција која је битна са различитих аспеката као што су адекватност и интерпретације.

Веома битан проблем у одређивању мере сличности је агрегација (фузија) више атрибута у један глобални

репрезентативни аспект - меру сличности. У постојећој пракси метод суме тежинских коефицијената парцијалних аспеката је најчешће коришћена агрегациона техника. Овакв приступ је адитиван и за све случајеве који нису адитивни је неадекватан. На пример, коришћење тежинске суме за агрегацију чак и у случају са само два атрибута ( $a, b$ ), не дозвољава реализацију природне потребе за захтевом у коме су важна оба атрибута заједно. У заједници која проучава мулти-атрибутивни метод доношења одлука овај проблем је препознат и као решење је употребљена теорија капацитивности, позната у «фази-заједници» као *фази мера* и *фази интеграл* [1].

У оваквом приступу адитивност је «релаксирана» монононошћу, за коју адитивност представља специјалан случај. Као последица тога, могућ спектар примене овог приступа је значајно шири. Али са логичког аспекта монононост је још увек снажан ограничавајући фактор јер многе логичке функције по својој природи нису монотоне. Интерполативна реализација Булове алгебре (ИБА) укључује све логичке функције и све интерполативне операторе (операторе генерализованих производа) [1].

Логичка агрегација као адекватан алат за агрегацију у општем случају, као и у случају управљања квалитетом је заснована на ИБА. ИБА је технички базирана на генерализованим Буловим полиномима (ГБП).

#### 1.1. Генерализовани псеудо-Булов полином [1]

Сваком елементу ИБА (логичкој функцији) одговара генерализовани Булов полином са способношћу да процесира све вредности примарних променљивих из јединичног реалног интервала  $[0, 1]$ . Интерполативни псеудо-Булов полином је конвексна линеарна комбинација посматраних елемената ИБА – генерализованих Булових полинома:

$$\pi\varphi^\circ(\|a_1\|, \dots, \|a_n\|) = \sum_{i=1}^m w_i \varphi_i^\circ(\|a_1\|, \dots, \|a_n\|), \quad (1)$$

$$\sum_{i=1}^m w_i = 1, w_i \geq 0, i = 1, \dots, m.$$

Из дефиниције генерализованих Булових полинома, интерполативни псеудо-Булов полином је дат у следећем изразу:

$$\pi\varphi_\mu^\circ(\|a_1\|, \dots, \|a_n\|) = \sum_{i=1}^m w_i \sum_{S \in \mathcal{P}(\Omega)} \chi_{\sigma(\varphi_i)}(S) \sum_{C \in \mathcal{P}(\Omega/S)} (-1)^{|C|} \bigotimes_{a_i \in S \cup C} \|a_i\|, \quad (1.1)$$

$$= \sum_{S \in \mathcal{P}(\Omega)} \mu(S) \sum_{C \in \mathcal{P}(\Omega/S)} (-1)^{|C|} \bigotimes_{a_i \in S \cup C} \|a_i\|$$

Структурна функција  $\mu$  интерполативног псеудо-Буловог полинома  $\pi\varphi_\mu^\circ$  је скуп функција

$$\mu: \mathcal{P}(\Omega) \rightarrow [0, 1], \quad \Omega = \{a_1, \dots, a_n\}$$

дефинисана следећим изразом:

$$\mu(S) = \sum_{i=1}^m w_i \chi_{\sigma(\varphi_i)}(S), \quad S \in P(\Omega), \quad \varphi_i \in BA(\Omega),$$

$$\sum_{i=1}^m w_i = 1, \quad w_i \geq 0, \quad i = 1, \dots, m$$

Где:  $\chi_{\sigma(\varphi_i)}$ ,  $i=1, \dots, m$  су логичке структурне функције одговарајућих Булових функција  $\varphi_i \in BA(\Omega)$ ,  $i=1, \dots, m$ .

Карактеристике псеудо-Буловог полинома зависе од генерализованог производа и његове структурне функције. Структурне функције могу бити класификоване као:

(а) адитивне, (б) монотоне и (с) генерализоване ( $(a) \subset (b) \subset (c)$ ).

### 1. 2. Логичка агрегација [1]

Полазна тачка је скуп примарних атрибута квалитета  $\Omega = \{a_1, \dots, a_n\}$ . Задатак логичке агрегације (ЛА) је фузија вредности парцијалних мера сличности у једну резултујућу глобалну репрезентативну вредност користећи логичке алате. И општем случају ЛА има два корака:

(1) Нормализација вредности примарних атрибута:

$$\| \cdot \|: \Omega \rightarrow [0,1]$$

Резултат нормализације је генерализована логички и/или  $[0, 1]$  вредност посматраног примарног атрибута, и

(2) Агрегација нормираних вредности примарних атрибута у једну резултујућу вредност помоћу псеудо-логичке функције као логичког агрегационог оператора:

$$Aggr: [0,1]^n \rightarrow [0,1]$$

Булова логичка функција  $\varphi$  је трансформисана у одговарајући генерализовани Булов полином (ГБП),  $\varphi^{\otimes}: [0, 1]^n \rightarrow [0, 1]$ . Заправо, сваком елементу Булове алгебре (над скупом примарних атрибута)  $\varphi_i \in BA(\Omega)$  одговара јединствени ГБП  $\varphi_i^{\otimes}(\|a_1\|, \dots, \|a_n\|)$ .

Псеудо-логичка функција је конвексна линеарна комбинација генерализованих Булових полинома дефинисаних изразима (1) и/или (1. 1).

Оператор логичке агрегације у општем случају је псеудо-логичка функција:

$$Agg_{\mu}^{\otimes}(\|a_1\|, \dots, \|a_n\|) = \pi \varphi_{\mu}^{\otimes}(\|a_1\|, \dots, \|a_n\|) \quad (2)$$

или

$$Agg_{\mu}^{\otimes}(\|a_1\|, \dots, \|a_n\|) = \sum_{S \in P(\Omega)} \mu(S) \sum_{C \in P(\Omega/S)} (-1)^{|C|} \otimes_{a_i \in S \cup C} \|a_i\|. \quad (2.1)$$

Мера агрегације је структурна функција псеудо-логичке функције – логички агрегациони оператор. Дакле, мера агрегације је скуп функција  $\mu: P(\Omega) \rightarrow [0, 1]$ , што у општем случају не мора бити монотона функција (генерализована капацитивност), дефинисана следећим изразом (3):

$$\mu(S) = \sum_{i=1}^m w_i \sigma_{\varphi_i}(S), \quad S \in P(\Omega), \quad \varphi_i \in BA(\Omega)$$

$$\sum_{i=1}^m w_i = 1, \quad w_i \geq 0, \quad i = 1, \dots, m \quad (3)$$

Као последица, логички агрегациони оператор зависи од одабраних: (а) мере агрегације и (б) оператора генерализованог производа. Помоћу одговарајућег избора мере агрегације  $\mu$  и генерализованог производа  $\otimes$  познати агрегациони оператори могу бити добијени као специјални случајеви.

Избор генерализованог производа зависи од конкретне природе проблема. На пример, ако су атрибути високо (позитивно или негативно) корелисани, добар избор за генерализовани производ је *min* функција, док је за случај слабе корелације атрибута добар избор генерализованог производа обичан производ.

## 2. ОПИС ОПШТЕГ КОНЦЕПТА МОДЕЛА ЗАКЉУЧИВАЊА НА ОСНОВУ СЛУЧАЈА

Један занимљив концепт пословне интелигенције који користи податке из прошлости (меморију) ради доношења одлука је закључивање на основу случајева (ЗОС).

ЗОС (*CBR - Case Based Reasoning*) је концепт који решава нове проблеме користећи решења старих проблема. Може да се каже да ЗОС решава проблеме користећи или адаптирајући решења старих проблема.

Процес ЗОС тече следећим током: јавља се нови проблем који треба решити. Доносилац одлуке тада има могућност да претражује тзв. базу знања да би нашао решења која су слична постојећем проблему. Затим се старо решење прилагођава новом проблему и генерише се решење.

Ако се два пута решава исти проблем у истом контексту (сви параметри су исти или “довољно” слични), памти се решење које на бољи начин решава проблем. Овим се постиже унапређивање процеса одлучивања у конкретном пословном процесу.

ЗОС има одређене претпоставке, које су од интереса за овај рад:

1. Универзум функционише по принципу реда, а не по принципу хаоса;
2. Ситуације у животу (проблеми) се понављају, па их вреди памтити и
3. Слични проблеми имају најчешће слична решења.

ЗОС се састоји из шест активности, које се зову шест RE (*Retrieve, Reuse, Revise, Review, Retain & Refine*). То су следеће активности [3]:

1. Проналажење знања (случаја) који одговара проблему;
2. Коришћење дела пронађеног знања за решавање проблема;
3. Прилагођавање, ако је потребно, дела пронађеног знања новом проблему;

4. Проверавање да ли је ново решење вредно памћења;
5. Памћење новог знања, уколико је тако одлучено у активности 4;
6. Прерада „новог” знања у бази знања, уколико за тим има потребе.

ЗОС се у односу на друге методе више ослања на људски фактор. ЗОС није самосталан у раду као неке друге технике за подршку одлучивању. Да би се овај процес вршио како треба, потребно је да доносилац одлуке поседује одређена знања из конкретнoг домена.

Шест активности ЗОС могу да се групишу у четири фазе одлучивања, сличних фазама других метода (вишеатрибутивно одлучивање - ВАО, вештачке неуронске мреже - ВНМ и Data Mining - DM). Те фазе су:

1. Структурирање проблема,
2. “Нормализација” података,
3. Агрегација података и
4. Анализа.

### 2. 1. Структурирање проблема [4]

Као и код осталих концепата пословне интелигенције, и код ЗОС је потребно да се проблем уреди, да се структурира. Код ЗОС се проблем уређује преко тзв. случајева. Пошто је проблематично цео случај памтити као велику количину неуређеног текста, проблем се раставља на подпроблеме, тј. критеријуме. Тиме случај добија своју структуру.

Случај је један од начина представљања знања. Сваки случај садржи критеријуме који су подељени на улазне и излазне критеријуме. *Улазни критеријуми* памте проблем, док *излазни критеријуми* памте решење проблема. Случај се, према томе, састоји од скупа критеријума чије вредности носе информације о проблему и скупа критеријума који носе информације о његовом решењу. У вредностима критеријума могу бити смештени, између осталог, и текстуални записи, фотографије, видео записи итд. Случај има особину семантичког богатства.

Подскуп скупа критеријума код случаја, који служе за претраживање базе знања, су индекси. Сваки случај се састоји из тзв. индексираних и неиндексираних критеријума. Индексирани критеријуми служе за претраживање базе знања. Неиндексирани критеријуми служе за описивање случаја и могу утицати на решење проблема уколико индексирани критеријуми не садрже довољно информација. Неиндексирани критеријуми не служе за упоређивање случајева (нису нормализовани), већ у себи носе семантичке информације неприкладне за нормализацију (нпр. мултимедијални подаци).

Индекси, тј. критеријуми за претраживање, би требало да буду: разумљиви, сврсисходни, довољно апстрактни (ради евентуалног проширивања базе случајева) и конкретни (ради прецизности претраживања).

Индексирање, тј. избор индекса, може да се ради ручно, али и аутоматски. Ручна метода тражи експертско знање, док аутоматска метода користи неке од техника DM, ВАО или ВНМ.

Иако постоје многобројне технике за аутоматско индексирање, некада је боље оставити доносиоцу одлуке да сам уради индексирање, уколико се радио о експерту у области који има прецизну представу шта жели да постигне, тј. када је способан да моделира решење.

Након што је урађено структурирање проблема, прелази се на део који треба да омогући прелаз из аналитичке у синтетичку фазу одлучивања, тј. нормализацију.

### 2. 2. Нормализација случајева [4]

Након што је проблем (случај) структуриран, потребно је све податке довести у облик који омогућава ефикасну манипулацију и управљање истима.

Случајеви имају обележје семантичког богатства чувања знања. Они треба да служе као преносиоци одређене приче, искуства о одређеном догађају из кога може нешто да се научи. Са друге стране, треба имати меру, када се прави база случајева (знања), између семантичког богатства одређеног случаја и ефикасног претраживања базе знања. Та мера се постиже дефинисањем индекса, део претходно описане фазе.

Код нормализације је потребно све случајеве организовати тако да све вредности буду упоредиве. Најчешће се у ту сврху користе познате нормализације преко метрика. L1 (Манхетн), L2 (векторска - Еуклидска), L $\infty$  (Чебишевљева) итд. Врста нормализације има бесконачно много, и није увек јасно зашто је боље користити некад једну, а не другу метрику, већ се искуственом провером долази до сазнања о употребљивости метрика.

Треба напоменути да нормализација може да има и шире значење. Нормализацију треба посматрати као процес у којем се одређеном скупу података додељује заједничко обележје у циљу лакшег управљања подацима.

### 2. 3. Агрегација података [4]

Након што је извршена анализа система и извршена нормализација, могуће је извршити синтезу података, тј. агрегацију.

Код ЗОС се користи концепт сличности (удаљености) којим се рачуна одстојање новог проблема од старих проблема. То значи да се све агрегације које се користе код осталих концепата пословне интелигенције (ПИ) могу користити и код ЗОС уз мале модификације.

Алгоритам за рачунање сличности може бити такав да користи погодности меморијске организације случајева. Насупрот претраживањима база података, које користи SQL језик и добија прецизне *crisp* информације, проналажење случајева користи понекад и хеуристике које проналазе фази (*fuzzy*) сличне вредности, с обзиром да, у општем случају, не постоји случај који прецизно (*crisp*) налази решење новог проблема.

Најпознатије методе које се користе за проналажење случајева су: метода најближег суседа, индукција, инду-

кција вођена знањем и откривање патерна. Ове методе могу бити самосталне или комбиноване у хибридне стратегије проналажења. Неко може да се запита зашто се овде уопште говори о агрегацији. Разлог томе јесте што све методе претраживања претварају више индекса у једну вредност (колону), на основу којих се врши сортирање и избор најближег случаја. Следи приказ најпознатијих метода за претраживање базе случајева (базе знања).

### Метода најближег суседа

Метода најближег суседа представља најчешће коришћену методу за рачунање близине између новог проблема и старих проблема. Ова метода одређује сличност постојећег случаја са новим случајем на основу тежинског збира индекса. Највећи проблем код ове методе је одређивање тежинских коефицијената. Проблем тежина (пондера) је највећи проблем науке о одлучивању уопште. Појава да мале промене тежина могу имати велики утицај на решење, чини одређивање пондера још значајнијим. За дефинисање пондера користе се све расположиве технике одлучивања DM, BHM, BAO (АХП), статистика итд. Често се консултује и експерт из области, али и даље проблем тежина остаје критична тачка науке о одлучивању.

Функција сличности се уобичајено представља преко Еуклидске норме. Наравно, Еуклидова норма представља само један од начина на који је могуће мерити сличност. Постоје бесконачно много начина на који може да се рачуна сличност, а што зависи од ситуације (контекста) у коме се примењује.

Технике адаптације код ЗОС могу да помогну да се сагледа да ли је случај добро структуриран, да ли је извршена квалитетна нормализација, да ли је агрегација урађена коректно. Додатно омогућава да се постојеће база случајева ефикасније искористи за решавање новог проблема. Оваквим начином рада, уз коришћење ИВА, систем ЗОС напредује током времена и постаје кориснији.

Патерни су један од најбољих начина чувања знања, мада неки аутори теорије патерна сматрају да је ово тек њихово секундарно својство.

### 3. РАЗВОЈ МОДЕЛА ЗА АНТИЦИПАЦИЈУ ЦЕНА НЕКРЕТНИНА

Могућности модела логичке агрегације проширују оквире постојећих фази модела. У досадашњим моделима коришћени су само тривијални атрибути, без интеракције између њих (без употребе логичких функција). Такође, за оператор агрегације коришћен је само обичан производ.

Модел логичке агрегације је генерализован и омогућава употребу логичких функција између атрибута (из коначног скупа могућих функција), као и употребу различитих оператора агрегације, као и различитих генерализованих производа (подкласа Т-норми уз испуњен аксиом ненегативности) [2].

Имплементација логичке агрегације је могућа у концепту ЗОС и то у фази агрегације података - одређивању мере сличности.

Задатак модела је процена цене некретне. Проблем је структуриран у виду познатих случајева некретнина и њихових цена, тј. некретнина са познатим ценама и нормираним вредностима атрибута организованим у базу знања. Улазни параметри модела су нормиране вредности атрибута некретнине чија се цена жели предвидети/проценити. Модел затим проналази у бази знања најсличнију некретнину користећи у овом раду разрађени математички апарат и предвиђа цену некретнине по основу цене најсличније некретнине из базе знања. Циљ овог рада је упоредна анализа резултата коришћења Еуклидске норме и ИБА као одговарајућих оператора агрегације који процесирају на вредносном нивоу и одређују меру сличности. Нормирање вредности атрибута ће бити описано у посебном одељку. Подразумева се употреба истог типа нормирања података из базе знања и улазних података.

#### 3. 1. Модел заснован на концептима ЗОС и ЛА

Базу знања представља матрица  $O$  коју чине постојеће некретнине-прототипови (објекти). Случајеви се састоје од скупа атрибута чије су вредности носиоци информације о проблему и одговарајуће јединичне цене која је носилац информације о решењу проблема. Сваки прототип  $O^i$  представља *случај* - вектор који чине нормиране вредности атрибута  $a_i^j \in [0, 1]$ , јединична цена  $c^j$  и површина некретнине  $p^j$ . Индекс случаја је  $j$ . Атрибути су дефинисани као:

$$\exists O_j \forall a_i^j, i \in (1, \dots, k)$$

За одређивање цене неке некретнине  $O^x$  потребно је нормирати вредности атрибута на исти начин као као што је учињено у бази знања. Затим се врши одређивање мере сличности  $\mu^j$  са некретнинама из базе знања.

Мера сличности је агрегирана величина. Добија се као резултат агрегације појединачних (парцијалних) мера сличности по сваком од атрибута понаособ.

Парцијална мера сличности је заправо мера логичке еквиваленције. Овакав избор агрегације парцијалних мера сличности је последица велике квалитативне сличности - ради се о поређењу некретнина по истој особини (атрибуту) и отвара се могућност употребе *min* функције за генерализовани производ, што је адекватно датом проблему. Одговарајући оператор агрегације вредности мере сличности је:

$$\mu^j = (a_i^j \Leftrightarrow a_i^x)^\otimes = [(a_i^j \cap a_i^x) \cup (Ca_i^j \cap Ca_i^x)]^\otimes = 1 - a_i^j - a_i^x + 2a_i^j \otimes a_i^x$$

$\otimes := \min$

Пошто се ради о агрегацији парцијалних мера сличности по истом атрибуту, за генерализовани производ користимо *min* функцију. Оператор агрегације врши пресликавање:

$$Aggr: [0, 1]^2 \rightarrow [0, 1]$$

Укупна мера сличности  $\mu^j$  се добија агрегацијом парцијалних мера сличности (по атрибутима):

$$\mu^j = \left( \bigcap_{i=1}^k \mu_i^j \right)^\otimes = \prod_{i=1}^k \mu_i^j, \quad \otimes = *$$

Пошто се ради о агрегацији по различитим атрибутима (занемарљива квалитативна сличност атрибута), за генерализовани производ користимо обичан производ. Оператор агрегације врши пресликавање:

$$Aggr: [0, 1]^j \rightarrow [0, 1]$$

Постоји могућност доделе тежинских коефицијената  $w_i$  парцијалним мерама сличности #f16. Одговарајућа укупна мера сличности је онда одређена изразом:

$$\mu^j = \sum_{i=1}^k w_i \mu_i^j, \\ \sum_{i=1}^k w_i = 1, w_i \geq 0, i = 1, \dots, k$$

Остаје отворена могућност и агрегације хибридног-мешовитог типа према потреби уз коришћење интерполативних псеудо-Булових полинома. Најсличнији прототип  $O^j \in O$  са  $O^x$  је онај са највећом вредношћу мере сличности  $\mu^j$ :

$$\exists O_x \forall c^j \in O^j, O^j \in O \text{ где је } j \text{ одабрано за } \mu^j = \max$$

На овај начин узима се вредност јединичне цене некретнине (случаја) из базе знања који је најсличнији улазном случају чију јединичну цену тражимо. Затим се вредност тако одређене јединичне цене придружује случају чија се јединична цена тражи.

Производ јединичне цене  $c^j$  и површине некретнине  $p^x$  даје тражену цену некретнине  $O^x$ .

$$C^x = c^j * p^x$$

### 3. 1. 2. Агрегација преко Еуклидске (L2) норме

Сваки од случаја  $O^j$  из базе знања  $O$  се може представити као тачка у  $n$ -димензионом реалном векторском простору, где је кардиналност базе векторског простора  $n$  одређена са бројем атрибута  $a_i^j, i = 1, \dots, k \Rightarrow n = k$ , док је удаљеност правоугле пројекције тачке од координатног почетка по свакој од оса једнака нормираним вредностима сваког од атрибута. На исти начин се могу представити и улазни подаци случаја  $O^x$ .

При коришћењу Еуклидске норме као оператора агрегације мере сличности заправо се врши рачун мере растојања тачака  $O^x$  и  $O^j$  представљених на претходно описан начин. Растојање је дефинисано са:

$$r^j = \left[ \sum_{i=1}^k (a_i^j - a_i^x)^2 \right]^{\frac{1}{2}}, \text{ где је } k \text{ број атрибута}$$

Такође, остаје и могућност доделе тежинских коефицијената сваком од квадрата разлике вредности атрибута  $(a_i^j - a_i^x)^2$ . Тада је растојање дефинисано са:

$$r^j = \left[ \sum_{i=1}^k (a_i^j - a_i^x)^2 w_i \right]^{\frac{1}{2}},$$

где је:  $\sum_{i=1}^k w_i = 1, w_i \geq 0, i = 1, \dots, k$

Уколико доделимо тежинске коефицијенте, вредности излаз Еуклидске норме задржава се у интервалу  $[0, 1]$ , при подразумеваним нормирањем улазних параметара такође на овај интервал. Тада је вредност укупне мере сличности једнака вредности логичке негације растојања.

$$\mu^j = (-r^j)^\otimes = 1 - r^j$$

Најсличнији прототип  $O^j \in O$  са  $O^x$  је онај са највећом вредношћу мере сличности  $\mu^j$ , док се у случају коришћења Еуклидске норме без тежинских коефицијената за највећу вредност мере сличности узима случај где је растојање тачака  $r_i^j$  најмање:

$$\mu^j = \max \Leftrightarrow r^j = \min$$

$$\exists O_x \forall c^j \in O^j, O^j \in O \text{ где је } j \text{ одабрано за } \mu^j = \max$$

Производ јединичне цене  $c^j$  и површине некретнине  $p^x$  даје тражену цену некретнине  $O^x$ .

$$C^x = c^j * p^x$$

### 3. 2. Квалитативно дефинисање атрибута и нормирање вредности

Некретнине у бази знања су станови са ширег подручја Београда. У складу са доступношћу информација, одабрани атрибути унети у базу знања су:

- макролокација (крај града или општина)
- структура (број просторија - соба)
- микролокација (спрат као један од микролокацијских параметара)
- лифт (постоји или не, микролокацијска особина)
- тип грејања (централно, етажно или ТА-термо акумулациона пећ)

Свака некретнина  $O^j$  унета у базу знања је описана наведеним атрибутима  $a_i^j$ , јединичном ценом  $c^j$  (EUR/m<sup>2</sup>), као и површином  $p^j$  (m<sup>2</sup>).

Нормирање атрибута *макролокација* (крај града или општина)  $a_1^j$

Први корак претварања назива локација у нумерички облик извршен је тражењем средње вредности јединичне цене  $c^j$  за сваку од локација које се појављују у бази знања. Затим следи други корак - постављање у монотони поредак (растући). Завршни корак је линеарна пројекција (скалирање) на интервал [0, 1]. На овај начин су вредности атрибута прилагођени процесирању помоћу математичког модела.

Овакав поступак нормирања вредности показује предности приликом појаве некретности чију цену желимо предвидети помоћу модела, а у бази знања не постоји некретности са њеном локацијом. Онда се може извршити субјективно поређење са локацијама из базе знања и доделити новој локацији вредност која је између две најсличније које постоје у бази знања (субјективно), узимајући у обзир удаљеност од центра града, удаљеност од других локација које постоје у бази знања, затим друге предности и мане те нове локације у складу са хедоничким приступом (тржишно препознате субјективне преференције које се одражавају на цену) и другим доступним релевантним информацијама.

Најмања просечна локацијска јединична цена  $\bar{c}^j$  у бази знања је 1087 EUR/m<sup>2</sup>, а највећа 2528 EUR/m<sup>2</sup>. Ради омогућавања уноса нових локација у базу, што може довести до промене просечних вредности за већ унете локације, као и појаву нових екстремних просечних вредности (ван досадашњег интервала ценовног распона), скалирање је извршено по принципу да је:

- за вредност 500 EUR/m<sup>2</sup> нормирана вредност 0,
- за вредност 3000 EUR/m<sup>2</sup> нормирана вредност 1.

Дакле, интервал максималног ценовног распона је релаксиран, проширен за по  $\approx 500$ . Вредности атрибута *макролокација* су нормиране по обрасцу:  $\frac{c^j - 500}{2500}$ .

МАКРОЛОКАЦИЈА	$\bar{c}^j$	НОРМИРАНЕ ВРЕДНОСТИ АТРИБУТА $a_1^j$
Периферија града *	1086. 92	0. 235
Раковица	1376. 92	0. 351
Церак	1600. 00	0. 440
Вождовац-Медак	1693. 40	0. 477
Чукарица	1747. 90	0. 499
Земун	1838. 26	0. 535
Звездара	1843. 04	0. 537
Вождовац	1898. 82	0. 560
Нови Београд	2046. 63	0. 619
Дорћол	2094. 11	0. 638
Дедиње	2382. 79	0. 753
Врачар	2404. 39	0. 762
Центар	2527. 95	0. 811

\* - периферија града овде подразумева Коњарник, Миљаковац, Ледине и Железник

Нормирање атрибута *структура* (број просторија - соба)  $a_2^j$

Овај атрибут узима вредности од 1 (за гарсоњере и једноспобне станове) до 5+ (више од пет соба). Нормирање је извршено скалирањем уз максимално искоришћење интервала од 0 до 1.

У случају унапређења базе знања уносом некретности са бројем соба који се означава нецелобројно (нпр. једноспобан стан), остављена је могућност нормирања на пола просторије тако што се нпр. двоиспобан стан нормира као средина између 2 и 3 (2,5), тј. атрибут узима нормирану вредност 0,3.

БРОЈ ПРОСТОРИЈА - СОБА	НОРМИРАНЕ ВРЕДНОСТИ АТРИБУТА $a_2^j$
1	0
2	0,2
3	0,4
4	0,6
5	0,8
више од 5	1

Нормирање атрибута *микролокација* (спрат као параметар микролокације)  $a_3^j$

Атрибут микролокација (спрат) узима вредности од -1 (ниско приземље или сутерен), до 8+ (више од 8). Нормирање је такође извршено скалирањем уз максимално искоришћење интервала од 0 до 1.

МИКРОЛОКАЦИЈА - СПРАТ	НОРМИРАНЕ ВРЕДНОСТИ АТРИБУТА $a_3^j$
-1	0
0	0,1
1	0,2
2	0,3
3	0,4
4	0,5
5	0,6
6	0,7
7	0,8
8	0,9
више од 8	1

Нормирање атрибута *лифт*  $a_4^j$

Нормирање вредности овог атрибута је тривијално. Уколико лифт постоји атрибут узима вредност 1, уколико не постоји, вредност је 0.

ЛИФТ	НОРМИРАНЕ ВРЕДНОСТИ АТРИБУТА $a_4^j$
постоји	1
не постоји	0

Нормирање атрибута *тип грејања*  $a_5^j$

У бази знања овај атрибут је приказан као низ знакова - стринг који описује тип грејања (централно, етажно, или

ТА). Нормиране вредности су додељене тако да у случају појаве типа грејања који се није раније појавио у бази знања, остаје могућност субјективне доделе вредности у складу са сличношћу са постојећим типовима грејања.

ТИП ГРЕЈАЊА	НОРМИРАНЕ ВРЕДНОСТИ АТРИБУТА $a_5^j$
етажно	0,9
централно	0,6
ТА	0,3

### 3. 3. Упоредни приказ резултата коришћења ЛА и Еуклидске норме

У бази знања се иницијално налазило 90 некретнина. Случајним избором је одабрано 10 које треба да тестирају модел (квалитет предвиђања). Одабраних 10 некретнина је уклоњено из базе знања и њихови параметри атрибута служили су као улазни подаци.

Одабране некретнине су:

ЛОКАЦИЈА	СТРУКТУРА	СПРАТ	ЛИФТ	ГРЕЈАЊЕ	ПОВРШИНА	ЦЕНА
ЦЕНТАР	ТРОСОБАН	1	НЕ	ЕТАЖНО	118	245,000
ВРАЧАР	ПЕТСОБАН	8	ДА	ЦЕНТРАЛНО	101	250,000
ЦЕНТАР	ТРОСОБАН	4	НЕ	ЕТАЖНО	88	249,000
ВРАЧАР	ДВОСОБАН	0	НЕ	ЦЕНТРАЛНО	54	120,000
ЗВЕЗДАРА	ДВОСОБАН	3	НЕ	ЦЕНТРАЛНО	42	65,000
ЗЕМУН	ДВОСОБАН	1	НЕ	ЦЕНТРАЛНО	58	98,500
ВОЈДОВАЦ-МЕДАК	ДВОСОБАН	1	НЕ	ЦЕНТРАЛНО	44	75,000
ЧУКАРИЦА	ГАРСОЊЕРА	2	НЕ	ТА	22	39,000
НОВИ БЕОГРАД	ДВОСОБАН	4	НЕ	ЦЕНТРАЛНО	66	137,000
ЦЕНТАР	ЧЕТВОРОСОБАН	4	НЕ	ЦЕНТРАЛНО	118	315,000

#### 3. 3. 1. Модел са логичком агрегацијом

Ове некретнине су дате са својим реалним ценама које ће служити за поређење са ценама које даје модел.

Пре улаза у модел заснован на логичкој агрегацији извршићемо логичку агрегацију два примарна атрибута у један сложени атрибут. Атрибути *спрат* и *лифт* могу се повезати логичком интеракцијом. Полазећи од теоријске претпоставке да уколико је некретнина у приземљу лифт небитан, а битност лифта расте са повећањем атрибута *спрат*, па је сличност јединичне цене очекивана, може се дефинисати нови сложени атрибут  $a_{34}^j$  који је дефинисан изразом:

$$a_{34}^j \circledast = (-a_3^j \vee a_4) \circledast = (1 - a_3^j) + a_4 - (1 - a_3^j) \otimes a_4 = 1 - a_3^j + a_3^j \otimes a_4, \quad \otimes := *$$

Након дефинисања овог сложеног атрибута, могуће је избацити из модела појединачне атрибуте *спрат* и *лифт*, јер су заступљени на нов начин преко сложеног атрибута.

На аритметичком (вредносном) нивоу, парцијална мера сличности по сваком од атрибута је одређена полиномом:

$$\mu_i^j = 1 - a_i^j - a_i^x + 2 \min(a_i^j, a_i^x), \quad i = 1, \dots, 4$$

Укупна мера сличности је дата изразом:

$$\mu^j = \sum_{i=1}^k w_i \mu_i^j, \quad k = 4,$$

тежински коефицијенти су респективно:

$$w_1 = 0,8, \quad w_2 = 0,05, \quad w_{3,4} = 0,1, \quad w_5 = 0,05$$

#### 3. 3. 2. Модел са Еуклидском нормом

Растојање је одређено изразом:

$$r^j = \left[ \sum_{i=1}^k (a_i^j - a_i^x)^2 w_i \right]^{\frac{1}{2}}$$

Мера сличности је:

$$\mu^j = 1 - r^j$$

Тежински коефицијенти атрибута су респективно:

$$w_1 = 0,8, \quad w_2 = 0,05, \quad w_3 = 0,05, \quad w_4 = 0,05, \quad w_5 = 0,05$$

Улазни параметри су идентични са параметрима модела заснованог на логичкој агрегацији.

#### Упоредни приказ резултата коришћења ЛА и Еуклидске норме

Поређење квалитета модела по основу избора оператора агрегације мере сличности биће приказано преко индекса мере одступања резултата модела од реалних вредности.

За дефиницију индекса мере одступања резултата  $\Delta$  од реалних вредности коришћена је дефиниција стандардне девијације, с тим што референца није средња вредност скупа дискретних података, већ реалне вредности цене сваке од некретнина.

$$\Delta = \sqrt{\sum_{i=1}^n (C_i^x - C_i)^2},$$

где су: број узорака  $n = 10$ ,  $C_i$  - реалне цене некретнина, и  $C_i^x$  - антиципиране цене (излаз модела).



МОДЕЛ (ЛОГИЧКА АГРЕГАЦИЈА)				МОДЕЛ (ЕУКЛИДСКА НОРМА)			
$C_i$	$C_i^*$	$ C_i^* - C_i $	$\Delta$	$C_i$	$C_i^*$	$ C_i^* - C_i $	$\Delta$
245,000	250,548	5,548	1,754. 41	245,000	273,623	28,623	9,051.45
250,000	256,385	6,385		250,000	202,000	48,000	
249,000	211,200	37,800		249,000	211,200	37,800	
120,000	121,935	1,935		120,000	121,935	1,935	
65,000	76,042	11,042		65,000	76,042	11,042	
98,500	103,796	5,296		98,500	98,864	364	
75,000	75,000	0		75,000	75,000	0	
39,000	36,474	2,526		39,000	36,474	2,526	
137,000	130,630	6,370		137,000	115,077	21,923	
315,000	283,200	31,800		315,000	282,605	32,395	

Модел са мањим индексом мере одступања резултата врши квалитетније одређивање/предвиђање цена некретнина. У овом раду (са ограниченим узорком) се показало да је модел који користи *логичку агрегацију* бољи антиципатор цена некретнина од модела који користи *Еуклидску норму* као оператор агрегације мере сличности.

### Закључак

Конвенционални алати за агрегацију су често неадекватни због ограничења у виду немогућности коришћења логичких интеракција између атрибута квалитета. Унапређивање модела коришћењем напредних метода логичке агрегације значајно проширује могућности прилагођавања модела специфичним потребама. Ово се нарочито огледа у могућности укључивања сложених логичких функција у којима се појављују нетривијални комплексни атрибути. Нови приступ третира логичке функције - парцијалне агрегационе захтеве, као генерализовани Булов полином који процесира вредности из јединичног реалног интервала  $[0, -1]$ . Агрегација у општем случају представља генерализовану псеудо-логичку функцију.

Такође, значајан фактор унапређења квалитета модела је адекватан избор генерализованог производа при вредносном (аритметичком) процесирању логичке агрегације.

С обзиром на недоступност потпуних информација које би довеле до квалитетних скупова података у бази знања, модели у овом раду имају ограничења по питању прецизности (неки атрибути који имају значајан фактор утицаја на цену некретнине не могу бити уграђени у базу знања). Такође, цене некретнина из базе знања нису сасвим

тржишно нивелисане, јер се не ради о ценама извршених купопродаја (ове цене нису доступне јавности), већ о ценама актуелне понуде.

Насупрот ограничењима, експериментисањем на доступном скупу података из базе знања и разним модификацијама модела (различит избор тежинских коефицијената, оператора агрегације и типова генерализованих производа), дошло се до резултата у виду два модела са солидним могућностима процене/предвиђања цена некретнина.

Модел заснован на логичкој агрегацији се показао бољим од модела који користи Еуклидску норму за процесирање агрегације мере сличности.

### Литература

- [1] MIRKOVIĆ M., HODOLIĆ J., RADOJEVIĆ D. - **AGGREGATION FOR QUALITY MANAGEMENT**, Yugoslav Journal of Operations Research 16 (2006), Number 2, 177-188
- [2] RADOJEVIĆ D. - **BOOLEAN FRAME IS ADEQUATE FOR TREATMENT OF GRADATION OR FUZZINESS EQUALLY AS FOR TWO-VALUED OR CLASSICAL CASE**, SISY 2006 • 4th Serbian-Hungarian Joint Symposium on Intelligent Systems
- [3] WATSON I., MARIR F. - **CASE-BASED REASONING: A REVIEW**, THE KNOWLEDGE ENGINEERING REVIEW, Vol. 9, No. 4, 1994, str 335-381
- [4] ДЕЛИБАШИЋ Б. - **ФОРМАЛИЗАЦИЈА ПРОЦЕСА ПОСЛОВНОГ ОДЛУЧИВАЊА ПРЕКО ПАТЕРНА**, докторска дисертација, ФОН 2007.
- [5] LIMSOMBUNCHAI V., GAN C., LEE M. - **HOUSE PRICE PREDICTION: HEDONIC PRICE MODEL VS. ARTIFICIAL NEURAL NETWORK**, American Journal of Applied Sciences 1 (3), 2004, str. 193-201
- [6] ZADEH L. - **FUZZY SETS** - Information and Control, No. 8, 1965, str. 338-353
- [7] ZADEH L., BELLMAN R. E. - **LOCAL AND FUZZY LOGICS, MODERN USES OF MULTIPLE-VALUED LOGIC** - J. M. Dunn and G. Epstein (eds.), Dordrecht: D. Reidel, 1977 str. 103-165
- [8] RADOJEVIĆ D. - **LOGICAL AGGREGATION BASED ON INTERPOLATIVE REALIZATION OF BOOLEAN ALGEBRA**, EUSFLAT Conf. (1) 2007. str. 119-126



Милан Мркаљ, докторант - Факултет Организационих Наука, Београд

Области интересовања: менаџмент, операциона истраживања, теорија одлучивања, вишекритеријумска оптимизација, математичко моделовање