

## Закључивање на основу случајева<sup>1</sup>

Постоје три претпоставке које се прихватају када се говори о закључивању на основу случајева (ЗОС):

1. Свет у коме живимо има у себи одређени ред, тј. не функционише по принципу хаоса.
2. Ситуације у животу се понављају и зато их вреди памтити и
3. Слични проблеми имају слична решења.

Случај представља уређен пар проблем-решење у одређеном контексту. Он представља решен проблем из прошлости. Памти се јер се претпоставља да може да се поново употреби у оном тренутку када се јави сличан проблем, решеном проблему из прошлости.

Случајеви се памте у базама података јер се побољшавањем решених случајева из прошлости кроз ново решавање проблема долази до облика знања који је најповољнији за одлучивање. Тај облик су патерни који се приказују уређеном тројком проблема, контекста, и решења. Они представљају најбољи могући начин како један проблем може да се реши. Проблем се код патерна решава уз минималан утрошак енергије, а решење је одређено природним законитостима. За разлику од случајева, они представљају одређену структуру на вишем ниво апстракције од случаја.

Упрошћено речено, када се један проблем много пута решава, долази се до знања које је потребно да се тај проблем реши на најбољи могући начин.

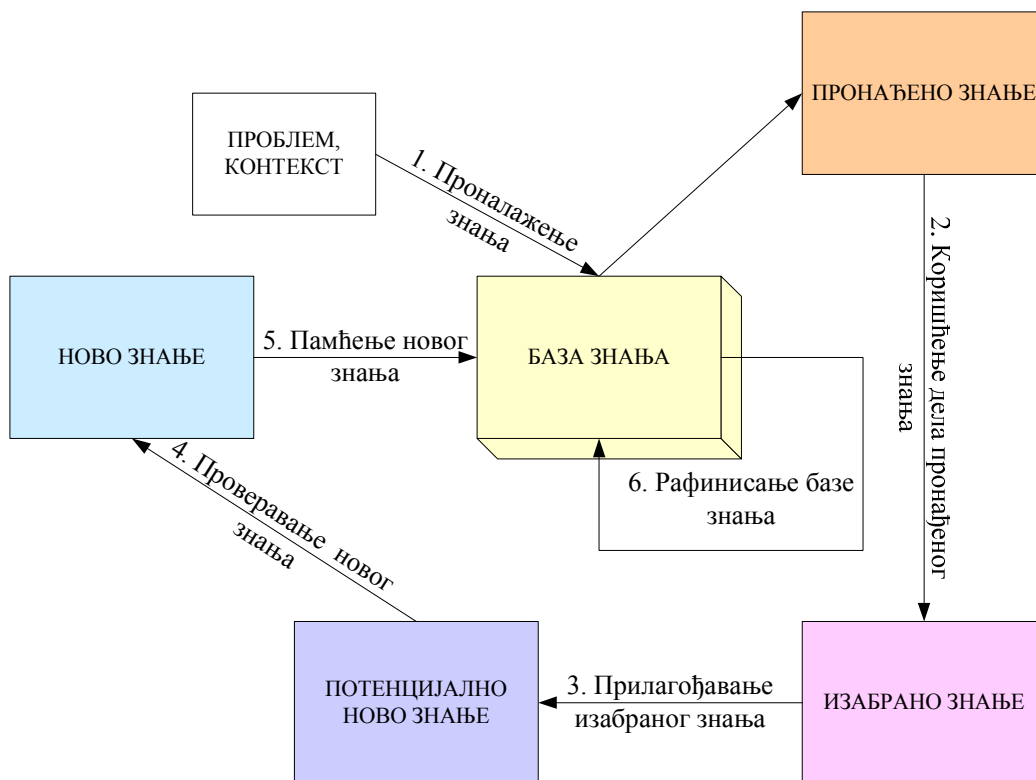
ЗОС је методологија која се састоји из неколико активности, који се зову шест RE<sup>2</sup>. То су следеће активности (Слика 6.1):

1. Проналажење знања (случаја) који одговара захтеву (проблему);
2. Коришћење дела пронађеног знања;
3. Прилагођавање, уколико има потребе, дела пронађеног знања новом захтеву;
4. Проверавање да ли је ново решење вредно памћења;
5. Памћење новог знања, уколико је тако одлучено под 4;
6. Рафинисање (прерада) новог знања у бази знања, уколико за тим има потребе.

---

<sup>1</sup> Са енглеског: Case-Based Reasoning

<sup>2</sup> Retrieve, Reuse, Revise, Review, Retain & Refine.



Слика 6.1. Животни циклус ЗОС

Битно је поменути да је ЗОС процес који умногоме зависи од људског фактора. ЗОС није самосталан у раду као неке друге технике за подршку одлучивању. Да би се овај процес вршио како треба, потребно је да и особа која ради поседује одређена знања из одговарајућег домена.

Најкорисније знање за одлучивање има управо форму случајева или, још боље, патерна. Знање представља информацију са додатком акције (решење проблема) и показатељем колико је случај из прошлости погодан за решавање новог проблема. Процес који говори колико је одређено знање корисно за решавање новог проблема зове се **валидација**. У наставку ће се показати како се формирају базе знања за одлучивање и како се врши процес валидације (одређивање колико је одређени случај погодан за решавање новог проблема).

База знања ће се изградити и чувати у редовима у Excel-у, док ће се валидација вршити преко методе најближих суседа и то конкретно преко Еуклидове норме. У једном реду у Excel-у може да се чува случај. Случај се састоји из више критеријума (колоне), од којих неки представљају проблем, док неки представљају решење.

Најчешће коришћена метода за валидацију случајева је метода најближег суседа. Формула за валидацију преко ове методе је:

$$\Phi (P, C) = \sum f (P_i, C_i) * T_i, i = (1, n)$$

**П** – нови проблем чије се решење тражи;

**С** – Случај који се налази у бази знања;

**ф (Пи, Си)** –сличност између и-те вредности критеријума случајева П и С;

**Ти** – тежина критеријума ( $\sum \text{Ти} = 1$ , врши се нормализација критеријума преко L1 метрике);

**н** – број критеријума<sup>3</sup>.

Сличност се обично израчунава преко Еуклидове норме. Пример за мерење удаљености по Еуклидовој норми могу бити две особе А (висина: 175цм, тежина 70кг) и особа Б (висина 185цм, тежина 90 кг). Удаљеност између ове две особе по Еуклидовој норми је  $((185-175)^2 + (90-70)^2)^{1/2} = 22,36$ . Наравно да постоје и бољи алгоритми за валидацију знања, али се они овде неће обрађивати.

Пример 1. Ако се прави база знања о продаји у једној пекари тада се проблем представља преко критеријума као што су: број потрошача, број пекара у близини, празник, временски услови, доба године, доба дана итд, док се решење представља преко критеријума: продаја белог хлеба, црног хлеба, кифли, бурека итд. у одређеном временском периоду. Код овог примера база знања се гради јер постоји жеља код власника пекаре да предвиди продају својих производа како би на најбољи могући начин користио своје ресурсе и нпр. могао да омогући да тачно зна када у току дана да производи одређене количине производа и да тако ти производи буду увек свежи.

У примеру који ће се показати у наставку доноси се одлука о давању кредита на скали од 1 до 5, где 1 представља оцену за недавање кредита, док 5 представља оцену за обавезно давање кредита.

База знања се састоји из петнаест случајева који се састоје од девет критеријума. Првих осам критеријума представљају проблем случаја, док оцена представља решење случаја.

Претпоставља се да долази нови корисник који тражи кредит. Његови подаци су респективно следећи: 3000, 2800, 0, 1, 2, 3, 4, 3

Сада је потребно видети који случај из базе знање највише одговара новом проблему. Да би се то видело, ради се валидација за свих петнаест случајева.

Редни број случаја	Месчни приход	Месечни трошак	Власник Стана?	Год. на тренут. послу	Год. на претх. послу	Год. на тренут. адреси	Год. на претх. адреси	Број издржав. лица	Оцена
1	3000	1500	0	2	8	6	2	5	3
2	850	425	1	3	3	25	25	1	3

<sup>3</sup> Битно је напоменути да не морају сви критеријуми из случаја да се користе за валидацију знања. Користе се само они критеријуми који добро представљају случај. Између осталог, случај може да садржи и слике, тонске записе итд.,

3	1000	3000	0	0.1	0.3	0.1	0.3	4	1
4	9000	2250	1	8	4	5	3	2	5
5	4000	1000	1	3	5	3	2	1	4
6	3500	2500	0	0.5	0.5	0.5	2	1	1
7	2200	1200	1	6	3	1	4	1	3
8	4500	3500	0	8	2	10	1	5	2
9	1200	1000	0	0.5	0.5	1	0.5	3	1
10	800	800	0	0.1	1	5	1	3	1
11	7500	3000	1	10	3	10	3	4	5
12	3000	1000	1	20	5	15	10	1	5
13	2500	700	1	10	5	15	5	3	5
14	3000	2600	1	6	1	3	4	2	2
15	7000	3700	1	10	4	10	1	4	4

Да би могло да се врши поређење, неопходно је да се уради нормализација базе података, као и новог случаја. У овом примеру користиће се  $L^\infty$  метрика. Норме  $L^\infty$  метрике су:

9000	3700	1	20	8	25	25	5
------	------	---	----	---	----	----	---

Нормализована табела одлучивања се добија дељењем колона базе података са одговарајућим нормама.

0.333333	0.405405	0	0.1	1	0.24	0.08	1
0.094444	0.114865	1	0.15	0.375	1	1	0.2
0.111111	0.810811	0	0.005	0.0375	0.004	0.012	0.8
1	0.608108	1	0.4	0.5	0.2	0.12	0.4
0.444444	0.27027	1	0.15	0.625	0.12	0.08	0.2
0.388889	0.675676	0	0.025	0.0625	0.02	0.08	0.2
0.244444	0.324324	1	0.3	0.375	0.04	0.16	0.2
0.5	0.945946	0	0.4	0.25	0.4	0.04	1
0.133333	0.27027	0	0.025	0.0625	0.04	0.02	0.6
0.088889	0.216216	0	0.005	0.125	0.2	0.04	0.6
0.833333	0.810811	1	0.5	0.375	0.4	0.12	0.8
0.333333	0.27027	1	1	0.625	0.6	0.4	0.2
0.277778	0.189189	1	0.5	0.625	0.6	0.2	0.6
0.333333	0.702703	1	0.3	0.125	0.12	0.16	0.4
0.777778	1	1	0.5	0.5	0.4	0.04	0.8

Нови случајеви се такође деле са нормама, при чему се њиховим вредностима додељује 1, чим имају вредност већу од своје норме.  $x_j = 1, s_j > norma_j$ .

Нормализована вредност новог случаја је:

0.333333	0.756756	0	0.05	0.25	0.12	0.16	0.6
----------	----------	---	------	------	------	------	-----

Како сви критеријуми нису подједнако битни за одлуку о давању кредита, потребно је увести пондере за критеријуме и тиме побољшати формулу за еуклидову норму.

Пондери се задају скалом од 1 до 9 (9-много утиче, 1-занемарљиво утиче). Пондери су респективно: 9, 9, 9, 3, 3, 2, 2, 5

Када се изврши нормализација L1 метриком (дељење са сумом свих пондера) добијају се нормализовани пондери: 0.21, 0.21, 0.21, 0.07, 0.07, 0.05, 0.05, 0.12.

Сада је могуће израчунати сличност старог случаја са новим случајем по формули за Еуклидску близину.

$$Blizina = \sqrt{\sum_j (x_{ij} - x_j)^2 * w_j}, \forall i \in I, \text{ где је } x_{ij} - \text{нормализована вредност елемента базе}$$

података која се налази на  $i$ -том реду и  $j$ -тој колони,  $x_j$  – нормализована вредност  $j$ -те колоне новог случаја  $w_j$  – тежина  $j$ -тог критеријума,  $i$  – ред алтернативе  $I$  – скуп алтернатива.

Добија се следећа колона:

0.2928
0.6352
0.1444
0.5711
0.5407
0.1567
0.5253
0.2139
0.2488
0.2760
0.5359
0.6052
0.5603
0.4700
0.5403

Да би се видело који случај највише одговара решењу проблема, врши се сортирање по близини. Сличност се добија када се од 1 одузме близина.

Редни број случаја	Оцена	Близина	Сличност
3	1	0.1444	0.8556
6	1	0.1567	0.8433
8	2	0.2139	0.7861
9	1	0.2488	0.7512
10	1	0.2760	0.7240
1	3	0.2928	0.7072
14	2	0.4700	0.5300
7	3	0.5253	0.4747
11	5	0.5359	0.4641

15	4	0.5403	0.4597
5	4	0.5407	0.4593
13	5	0.5603	0.4397
4	5	0.5711	0.4289
12	5	0.6052	0.3948
2	3	0.6352	0.3648

Примећује се да је најсличнији стари случај који може да реши нови проблем, случај 3 са оценом 1 и сличношћу од 85,56%, па доносимо одлуку да одбијемо кредит.

*Питања за студенте:*

1. Шта је закључивање на основу случајева?
2. Које су основне претпоставке код закључивања на основу случајева?
3. Шта су случајеви, а шта патерни?
4. Шта је знање, а шта валидација знања?

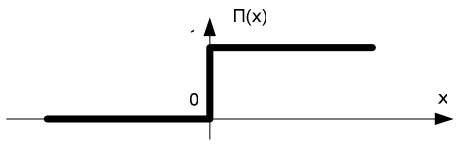
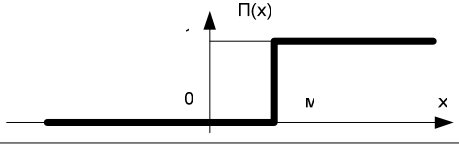
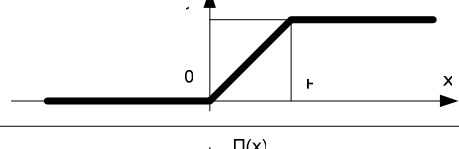
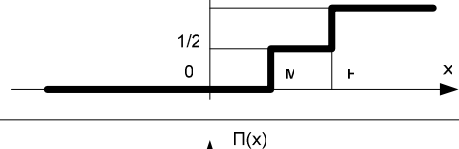
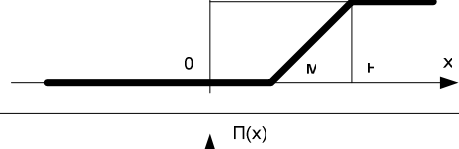
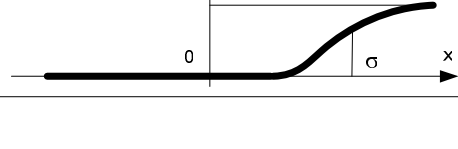
## Закључивање на основу случајева – други део

У поглављу шест је показана основна идеја и начин једноставне реализације система ЗОС. Највећи недостатак је представљао процес валидације преко Еуклидове норме, зато што се сви критеријуми посматрају на једнак начин, а једини начин да се изрази одређена преференција према критеријуму представљају пондери, што је за решавање практичних проблема, често, недовољно поуздано.

Да би се донекле исправио процес валидације, могуће је увести теорију преференције за претраживање базе знања и вредновање знања. Теорија преференције пружа више могућности при изражавању преференција ДО у тражењу одговарајућег случаја.

За претраживање базе знања користиће се шест типова критеријума из методе PROMETHEE 2. Графичка представа ових типова представљена је у табели 7.1.

Табела 7.1. Типови преференције код методе PROMETHEE 2

Функција преференције $\Pi(x)$	Врста општег критеријума	
	Тип 1. Обичан критеријум $\Pi(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases}$	
	Тип 2. Квази критеријум $\Pi(x) = \begin{cases} 0, & x \leq m \\ 1, & x > m \end{cases}$	$m$
	Тип 3. Линеарни критеријум $\Pi(x) = \begin{cases} 0, & x < c \\ x/n, & c \leq x \leq r \\ 1, & x > r \end{cases}$	$r$
	Тип 4. Ниво критеријум $\Pi(x) = \begin{cases} 0, & x \leq m \\ 1/2, & m < x < r \\ 1, & x \geq r \end{cases}$	$m, r$
	Тип 5. Индиферентни критеријум $\Pi(x) = \begin{cases} 0, & x \leq m \\ (x-m) / (r-m), & m < x \leq r \\ 1, & x > r \end{cases}$	$m, r$
	Тип 6. Гаусов критеријум $\Pi(x) = 1 - e^{-x^2 / 2\sigma^2}$	$\sigma$

Пројектанту базе знања могу да послуже следеће реченице да би се лакше снашао са употребом функција преференције<sup>4</sup>:

**Критеријум 1.** Две алтернативе су различите по одређеном критеријуму, ако између њих постоји и најмања разлика.

**Критеријум 2.** Две алтернативе су различите по одређеном критеријуму, ако су различите за више од М.

**Критеријум 3.** Две алтернативе су линеарно (сразмерно) различите по одређеном критеријуму ако је разлика између њих мања од Н, док су тотално различите ако је разлика већа или једнака Н.

**Критеријум 4.** Две алтернативе су једнаке уколико је разлика између њих мања или једнака М, упола различите (или сличне) уколико је разлика између њих између М и Н, а тотално различите уколико је разлика већа или једнака Н.

**Критеријум 5.** Две алтернативе су једнаке уколико је разлика између њих мања или једнака М, линеарно различите уколико је разлика између њих између М и Н, а тотално различите уколико је разлика већа од Н.

**Критеријум 6.** Разлика између две алтернативе се мери Гаусовим критеријумом, при чему је стандардна девијација  $\sigma$ .

Сада ће се показати како се теорија преференције користи за претраживање базе знања. Користиће се исти пример као у поглављу шест.

За претраживање по критеријумима изабрани су следећи типови критеријума, као и одговарајући параметри, респективно:

Тип критеријума	2	2	1	3	3	3	3	3
Параметар	500	400		5	5	6	6	5

Пондери за критеријуме су респективно:

Пондери	0.21	0.21	0.21	0.07	0.07	0.05	0.05	0.13
---------	------	------	------	------	------	------	------	------

Ако је случај за који се врши претраживање базе знања исти као у претходном примеру, а чува се у пољима од С26 до Ј26 (као на слици 7.2), а пондери у пољима С24 до Ј24, тада је за сваки критеријум за сваку алтернативу потребно рачунати удаљеност вредности од новог случаја (проблема).

Следе Excel формуле за прву алтернативу, по првом и трећем и четвртном критеријуму. За остале критеријуме се неће давати изглед формула, зато што је формула за други критеријум иста као и за први критеријум, а формула за пети, шести и седми критеријум је иста као формула за четврти критеријум.

<sup>4</sup> Марјановић М, Чупић М, Делибашевић Б и Сукновић М, *Пројектовање софтверске подршке за избор најповољније понуде у поступку набавке*, СИМОРГ 2004, 2004, стр. 6.



Формула за први критеријум, ако су испуњени услови као на слици 7.2, је  $=IF(ABS^5(C5-C\$26)>C\$22,1, 0)*\$C\$24$

	A	B	C	D	E	F	G	H	I	J	K
1											
2						Год. на	Год. на	Год. на	Год. на	Број	
3		Месечни	Месечни	Власник	тренут.	претх.	тренут.	претх.	издр.		
4		Рбр	приход	трошак	стана?	послу	послу	адреси	адреси	лица	Оцена
5		1	3000	1500	0	2	8	6	2	5	3
6		2	850	425	1	3	3	25	25	1	3
7		3	1000	3000	0	0.1	0.3	0.1	0.3	4	1
8		4	9000	2250	1	8	4	5	3	2	5
9		5	4000	1000	1	3	5	3	2	1	4
10		6	3500	2500	0	0.5	0.5	0.5	2	1	1
11		7	2200	1200	1	6	3	1	4	1	3
12		8	4500	3500	0	8	2	10	1	5	2
13		9	1200	1000	0	0.5	0.5	1	0.5	3	1
14		10	800	800	0	0.1	1	5	1	3	1
15		11	7500	3000	1	10	3	10	3	4	5
16		12	3000	1000	1	20	5	15	10	1	5
17		13	2500	700	1	10	5	15	5	3	5
18		14	3000	2600	1	6	1	3	4	2	2
19		15	7000	3700	1	10	4	10	1	4	4
20											
21		Тип.	2	2	1	3	3	3	3	3	
22		Пар.	500	400		5	5	6	6	5	
23											
24		Понд	0.21	0.21	0.21	0.07	0.07	0.05	0.05	0.13	1
25											
26		16	3000	2800	0	1	2	3	4	3	
27											

Слика 7.2. Изглед Excel табеле у којој се чува база знања

Формула за трећи критеријум је тада  $=IF(ABS(E5-E\$26)<>0,1, 0)*\$E\$24$ .

Формула за четврти критеријум је  $=IF(ABS(F5-F\$26)>=F\$22,1, ABS(F5-F\$26)/F\$22)*F\$24$

Потребно је добити следећу табелу удаљености алтернатива по свим критеријумима респективно од новог случаја. Ове вредности могу се памтити у пољима од N5 до U19.

	0	0.21	0	0.014	0.07	0.025	0.017	0.052
	0.21	0.21	0.21	0.028	0.01	0.05	0.05	0.052
	0.21	0	0	0.0126	0.02	0.024	0.031	0.026
	0.21	0.21	0.21	0.07	0.03	0.017	0.008	0.026
	0.21	0.21	0.21	0.028	0.04	0	0.017	0.052

<sup>5</sup> Код претраживања се користи разлика апсолутних вредности, јер за разлику од Терије одлучивања, критеријуми немају тип екстремизације (минимизација или максимизација)

0	0	0	0.007	0.02	0.021	0.017	0.052
0.21	0.21	0.21	0.07	0.01	0.017	0	0.052
0.21	0.21	0	0.07	0	0.05	0.025	0.052
0.21	0.21	0	0.007	0.02	0.017	0.029	0
0.21	0.21	0	0.0126	0.01	0.017	0.025	0
0.21	0	0.21	0.07	0.01	0.05	0.008	0.026
0	0.21	0.21	0.07	0.04	0.05	0.05	0.052
0	0.21	0.21	0.07	0.04	0.05	0.008	0
0	0	0.21	0.07	0.01	0	0	0.026
0.21	0.21	0.21	0.07	0.03	0.05	0.025	0.026

Удаљеност алтернатива од новог случаја се рачуна као збир редова горње табеле и износи:

Рбр	Удаљеност
1	0.38766667
2	0.824
3	0.3274
4	0.779
5	0.76866667
6	0.1175
7	0.78266667
8	0.617
9	0.49383333
10	0.48826667
11	0.58833333
12	0.684
13	0.59033333
14	0.32
15	0.829

Сличност се рачуна као 1-удаљеност и може да се изрази процентуално као у следећој табели:

Рбр	Близина
1	61.23%
2	17.60%
3	67.26%
4	22.10%
5	23.13%
6	88.25%
7	21.73%
8	38.30%
9	50.62%
10	51.17%
11	41.17%
12	31.60%

13	40.97%
14	68.00%
15	17.10%

Када се случајеви поређају по сличности добија се следећи поредак:

				Год. на	Год. на	Год. на	Год. на	Број		
	Месечни	Месечни	Власник	тренут.	претх.	тренут.	претх.	издр.		
Рбр	приход	трошак	стана?	послу	послу	адреси	адреси	лица	Оцена	Близина
6	3500	2500	0	0.5	0.5	0.5	2	1	1	88.25%
14	3000	2600	1	6	1	3	4	2	2	68.00%
3	1000	3000	0	0.1	0.3	0.1	0.3	4	1	67.26%
1	3000	1500	0	2	8	6	2	5	3	61.23%
10	800	800	0	0.1	1	5	1	3	1	51.17%
9	1200	1000	0	0.5	0.5	1	0.5	3	1	50.62%
11	7500	3000	1	10	3	10	3	4	5	41.17%
13	2500	700	1	10	5	15	5	3	5	40.97%
8	4500	3500	0	8	2	10	1	5	2	38.30%
12	3000	1000	1	20	5	15	10	1	5	31.60%
5	4000	1000	1	3	5	3	2	1	4	23.13%
4	9000	2250	1	8	4	5	3	2	5	22.10%
7	2200	1200	1	6	3	1	4	1	3	21.73%
2	850	425	1	3	3	25	25	1	3	17.60%
15	7000	3700	1	10	4	10	1	4	4	17.10%

Види се да је најближи случај новом случају, случај 6 са 88,25%.

*Питања за студенте:*

1. Зашто теорија преференције даје боље резултате у претраживању од еуклидове норме?
2. Зашто се код упоређивања алтернатива, по одређеном критеријуму, користи апсолутна разлика?