

# Алгоритмы построения нечетких классификаторов несбалансированных данных с отбором признаков на основе «гравитационного поиска» и «прыгающих лягушек»

Доклад аспиранта 3 года обучения  
Бардамовой Марины Борисовны

05.13.17 Теоретические основы  
информатики

Научный руководитель:  
профессор кафедры КИБЭВС, доктор технических  
наук, профессор  
Ходашинский Илья Александрович

# Несбалансированные данные

**Данные, в которых присутствует значительная разница между количеством экземпляров разных классов.**

Примеры:

- сетевой трафик;
- технологические процессы;
- медицинские диагностические данные;
- транзакции банковских операций,
- и тому подобное.

**Цель: улучшить качество нечеткой классификации при наличии дисбаланса в данных**

# Предлагаемая фитнес-функция 3

Для обеспечения сбалансированной точности при обучении классификатора предлагается объединить две метрики:

$$score = \gamma * GM + (1 - \gamma) * Acc$$

$\gamma$  – коэффициент приоритета между метриками,  
 $\gamma \in [0; 1]$

$GM$  – средняя  
геометрическая точность,  
 $Acc$  – общая точность

Если  $\gamma = 1$ , то  $score = GM$ ;  
Если  $\gamma = 0$ , то  $score = Acc$ .

# Структура нечеткого классификатора 4

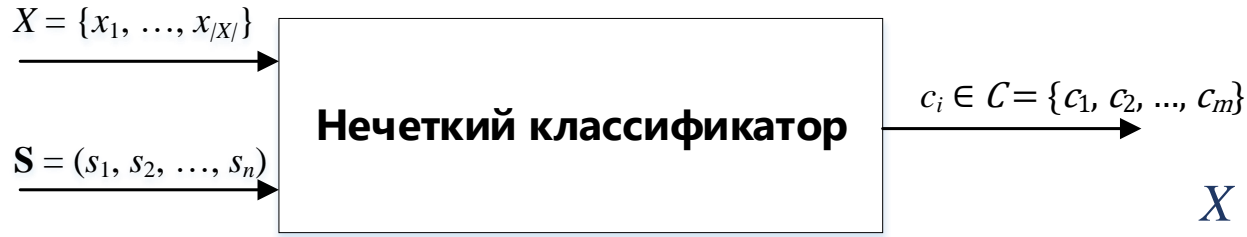


Рисунок 1 – Модель нечеткого классификатора

**$R_{ij}$  : ЕСЛИ  $s_1 \wedge x_1$  есть  $A_{1i}$  И  $s_2 \wedge x_2$  есть  $A_{2i}$   
И ... И  $s_n \wedge x_n$  есть  $A_{ni}$ , ТО class есть  $c_j$**

- $s_i = 1$  – признак участвует в классификации;
- $s_i = 0$  – признак не участвует в классификации.

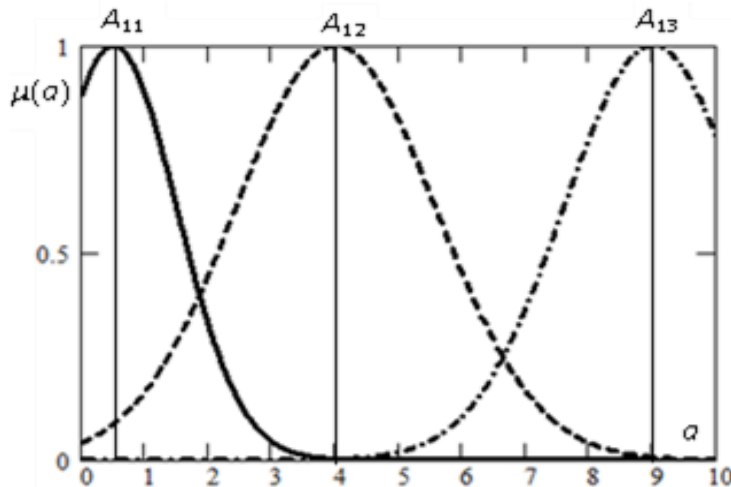


Рисунок 2 – Пример разбиения признака на три терма

$X = \{x_1, \dots, x_n\}$  –

МНОЖЕСТВО ВХОДНЫХ  
признаков;

$C = \{c_1, c_2, \dots, c_m\}$  –  
МНОЖЕСТВО МЕТОК КЛАССОВ;

$S = (s_1, s_2, \dots, s_n)$  –  
бинарный вектор  
признаков;

$A_{li}$  – нечеткий терм,  
характеризующий  $l$ -ый  
признак в  $i$ -ом правиле,

$(i \in [1, R])$

**1 Разработанный алгоритм формирования базы нечетких правил для несбалансированных данных позволяет создавать базы правил, демонстрирующие большую среднюю геометрическую точность по сравнению с алгоритмом экстремальных значений признаков классов.**

Соответствует пункту 5 паспорта специальности: Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечения.

Позволяет создать минимальное количество правил, равное числу классов.

**Достоинство:** классификатор обладает низкой вычислительной сложностью;

**Недостаток:** точность классификации может быть достаточно низкой даже после настройки параметров.

**Возможное решение:** дополнять базу правил с помощью метаэвристического алгоритма.

**Входные данные:** таблица обучения, база правил от первоначального алгоритма генерации структуры.

1. создание популяции векторов, содержащих дополнительное нечеткое правило;
2. построение классификаторов и оценка фитнес-функции;
3. сортировка векторов по убыванию фитнес-функции;
4. локальный поиск: изменение параметров термов нового правила согласно принципу алгоритма «прыгающих лягушек»;
5. возвращение к шагу 2 до достижения критерия остановки;
6. вывод базы правил с лучшим значением фитнес-функции.

$$fit(B \cup X) = error(B) - error(B \cup X)$$

$B$  – множество исходных правил,  
 $X$  – добавляемое правило.

## Ошибка классификации

$$error(Z) = 1 - (\gamma * GM(Z) + (1 - \gamma) * Acc(Z))$$

$GM(Z)$  – среднее геометрическое точности,  
 $Acc(Z)$  – общая точность,  
 $\gamma$  – коэффициент приоритета.



# Эксперимент 1: бинарные наборы

9

Сравнение результатов классификации при использовании только ЭК и комбинации ЭК с  $SFLA_{add}$ .

Data sets	Аббр-па	Признаки	$inst_{all}$	$inst+$	$inst-$	$IR$
wisconsin	wis	9	683	239	444	1,86
pima	pm	8	768	268	500	1,87
haberman	hbr	3	306	81	225	2,79
vehicle0	vhc0	18	846	199	647	3,25
newthyroid2	nwth2	5	215	35	180	5,14
segment0	sgm0	19	2308	329	1979	6,02
page-blocks0	pgbl0	10	5472	559	4913	8,79
vowel0	vwl0	13	988	90	898	9,98
cleveland-0vs4	clv04	13	177	13	164	12,62
ecoli4	ecl4	7	336	20	316	15,8
abalone9-18	abl9vs18	7/8	732	43	689	16,02
yeast4	yst4	8	1484	51	1433	28,1

# Усредненные результаты по 5 запускам 10

Критерий: средняя геометрическая точность

Алгоритм	ЭК	ЭК + SFLA <sub>add</sub>			
<i>R</i>	2	3	4	7	9
wisconsin	73,39	93,47 ± 1,04	<b>95,09</b> ± 1,10	94,95 ± 1,62	94,53 ± 1,01
pima	55,62	68,08 ± 1,90	70,57 ± 1,55	71,63 ± 1,77	<b>72,16</b> ± 2,32
haberman	44,28	46,05 ± 3,43	50,54 ± 3,81	54,98 ± 4,38	<b>56,45</b> ± 4,76
vehicle0	55,50	71,81 ± 4,19	75,50 ± 3,34	81,63 ± 1,84	<b>86,36</b> ± 2,18
newthyroid2	<b>99,16</b>	97,94 ± 0,47	98,18 ± 0,09	96,77 ± 2,54	96,83 ± 2,08
segment0	88,06	92,65 ± 1,09	94,02 ± 0,98	88,95 ± 1,04	<b>97,58</b> ± 0,82
page-blocks0	63,62	74,13 ± 1,78	81,85 ± 1,57	79,84 ± 2,26	<b>88,21</b> ± 0,85
vowel0	83,87	84,38 ± 1,02	86,62 ± 1,23	88,91 ± 3,10	<b>92,11</b> ± 3,37
cleveland-0vs4	<b>69,91</b>	67,89 ± 12,23	64,04 ± 14,61	49,22 ± 14,98	53,28 ± 23,38
ecoli4	68,70	87,98 ± 4,74	89,44 ± 3,10	<b>89,92</b> ± 3,11	84,32 ± 2,99
abalone9-18	58,73	67,44 ± 3,25	71,39 ± 3,72	<b>75,02</b> ± 2,30	72,24 ± 7,25
yeast4	65,74	70,16 ± 2,50	75,65 ± 2,74	<b>79,87</b> ± 2,13	77,26 ± 3,91
Среднее	68,88	76,83 ± 3,14	79,41 ± 3,15	79,31 ± 3,42	<b>80,94</b> ± 4,58

# Усредненные результаты по 5 запускам

Критерий: точность наименьшего класса

Алгоритм	ЭК	ЭК + SFLA <sub>add</sub>			
		3	4	7	9
<i>R</i>	2				
wisconsin	94,14	91,13 ± 1,58	94,73 ± 1,74	93,64 ± 2,79	92,96 ± 2,05
pima	57,48	68,54 ± 5,86	70,09 ± 5,24	70,53 ± 3,35	70,37 ± 3,62
haberman	54,34	34,24 ± 3,80	47,53 ± 7,90	48,31 ± 7,71	50,60 ± 7,29
vehicle0	34,72	69,43 ± 8,76	79,96 ± 9,84	84,42 ± 4,95	87,83 ± 3,10
newthyroid2	100,00	96,57 ± 0,91	97,14 ± 0,00	94,86 ± 5,03	94,86 ± 3,66
segment0	84,21	95,08 ± 1,65	94,59 ± 1,44	87,43 ± 2,91	96,97 ± 1,60
page-blocks0	42,76	62,40 ± 4,42	80,75 ± 4,02	76,82 ± 5,42	87,69 ± 1,50
vowel0	84,44	83,56 ± 1,42	87,11 ± 1,78	87,78 ± 5,60	89,56 ± 5,96
cleveland-0vs4	66,67	56,67 ± 17,60	51,33 ± 17,07	34,67 ± 12,80	40,67 ± 24,80
ecoli4	50,00	85,00 ± 10,40	86,00 ± 4,80	83,00 ± 5,60	73,00 ± 5,20
abalone9-18	50,56	66,06 ± 6,58	68,94 ± 5,82	70,61 ± 3,78	64,67 ± 12,31
yeast4	74,36	64,65 ± 5,40	75,42 ± 6,63	77,64 ± 3,91	70,73 ± 6,57

# Лучшие результаты среди 5 запусков

12

## Критерий: средняя геометрическая точность

Алгоритм	ЭК	ЭК + SFLA <sub>add</sub>				Прирост
		3	4	7	9	
<i>R</i>	2	3	4	7	9	
wisconsin	73,39	95,06	96,85	<b>97,17</b>	96,09	+ 23,78
pima	55,62	69,62	73,09	74,42	<b>75,42</b>	+ 19,80
haberman	44,28	50,76	55,03	61,88	<b>62,21</b>	+ 17,93
vehicle0	55,50	79,50	80,27	84,95	<b>89,50</b>	+ 34,00
newthyroid2	99,16	98,24	98,24	<b>99,72</b>	<b>99,72</b>	+ 0,56
segment0	88,06	94,40	95,62	91,55	<b>98,91</b>	+ 10,85
page-blocks0	63,62	76,49	84,49	84,20	<b>88,94</b>	+ 25,32
vowel0	83,87	85,64	88,19	93,12	<b>96,68</b>	+ 12,81
cleveland	69,91	84,36	83,74	72,87	<b>85,31</b>	+ 15,40
ecoli4	68,70	<b>96,28</b>	93,24	93,68	88,46	+ 27,58
abalone9-18	58,73	72,63	77,24	78,17	<b>83,94</b>	+ 25,21
yeast4	65,74	73,13	79,26	83,76	<b>83,79</b>	+ 18,05
<b>Среднее</b>	68,88	81,34	83,77	84,62	<b>87,41</b>	+ 19,27

## Критерий: средняя геометрическая точность

Алг.	Chi-3	Chi-5	Ishibuchi05	E-Algorithm	HFRBCS	C4.5	ЭК + SFLA <sub>add</sub>
wis	88,91 ± 2,13	43,58 ± 5,86	95,78 ± 1,38	96,01 ± 1,55	88,24 ± 1,63	95,44 ± 2,01	95,09 ± 1,1 (4)
pm	66,8 ± 5,93	66,78 ± 2,28	71,1 ± 4,45	55,01 ± 4,64	68,72 ± 5,26	71,26 ± 4,05	72,16 ± 2,32 (9)
hbr	58,91 ± 6,03	60,4 ± 2,4	62,65 ± 2,84	4,94 ± 11,06	57,08 ± 4,09	61,32 ± 3,85	56,45 ± 4,76 (9)
vhc0	86,41 ± 3,06	84,93 ± 1,61	75,94 ± 1,42	39,07 ± 16,49	88,92 ± 1,96	91,1 ± 2,7	86,36 ± 2,18 (9)
nwth2	89,81 ± 10,77	96,34 ± 6,65	94,21 ± 4,23	88,57 ± 3,82	99,72 ± 0,63	96,51 ± 4,87	98,18 ± 0,09 (4)
sgm0	94,9 ± 0,45	95,88 ± 1,21	42,47 ± 2,79	95,33 ± 1,14	97,51 ± 1,11	99,26 ± 0,61	97,58 ± 0,82 (9)
pgbl0	79,91 ± 4,29	87,25 ± 1,94	32,16 ± 9,61	64,51 ± 2,79	91,4 ± 0,67	94,84 ± 1,52	88,21 ± 0,85 (9)
vwl0	98,3 ± 0,61	97,8 ± 1,84	89,03 ± 6,63	89,63 ± 6,09	98,82 ± 1,62	94,74 ± 5,22	92,11 ± 3,37 (9)
ecl4	91,27 ± 7,43	92,11 ± 8,35	86,92 ± 8,65	92,43 ± 8,24	93,02 ± 8,17	81,28 ± 11,67	89,92 ± 3,11 (7)
abl9-18	63,93 ± 11	66,47 ± 10,67	65,78 ± 9,23	32,29 ± 20,61	67,56 ± 14,01	53,19 ± 8,25	75,02 ± 2,30 (7)
yst4	82,99 ± 3,1	83,07 ± 2,58	71,36 ± 23,29	32,16 ± 20,59	82,64 ± 2,29	65 ± 8,94	79,87 ± 2,13 (7)
Средн.	82,03 ± 4,98	79,52 ± 4,13	71,58 ± 6,77	62,72 ± 8,82	<b>84,88 ± 3,77</b>	82,18 ± 4,88	84,63 ± 2,09

Сравниваемые алгоритмы	p-value	STS	Нулевая гипотеза
Chi-3 - ЭК+SFLA <sub>add</sub>	0,213	-1,245	принять
Chi-5 - ЭК+SFLA <sub>add</sub>	0,594	-0,533	принять
Ishibuchi05 - ЭК+SFLA <sub>add</sub>	0,021	-2,312	откл.
E-Algorithm - ЭК+SFLA <sub>add</sub>	0,013	-2,49	откл.
HFRBCS - ЭК+SFLA <sub>add</sub>	0,79	0,267	принять
C4.5 - ЭК+SFLA <sub>add</sub>	0,859	-0,178	принять

Сравнение результатов классификации при использовании только ЭК и комбинации ЭК с  $SFLA_{add}$ .

Data sets	Классы	Признаки	$inst_{all}$	$IR$
penbased	10	16	1100 (115:114:114:106:114:106:105:115:105:106)	1.1
new-thyroid	3	5	215 (150:35:30)	5
dermatology	6	34	366 (112:61:72:49:52:20)	5.6
balance	3	4	625 (288:49:288)	5.88
thyroid	3	21	720 (17:37:666)	39.18

# Усредненные результаты по 5 запускам

16

Набор данных		АЭПК	АЭПК + SFLA <sub>add</sub>			
penbased	<i>R</i>	10	11	12	15	17
	<i>GM</i>	58,35	59,63 ± 1,14	61,48 ± 1,16	64,79 ± 1,52	<b>67,13 ± 1,58</b>
newthyroid	<i>R</i>	3	4	5	8	10
	<i>GM</i>	90,20	<b>92,46 ± 1,37</b>	92,24 ± 2,06	91,19 ± 2,78	89,02 ± 2,96
dermatology	<i>R</i>	6	7	8	11	13
	<i>GM</i>	85,92	<b>86,24 ± 0,45</b>	83,28 ± 6,03	84,68 ± 2,76	86,22 ± 2,18
balance	<i>R</i>	3	4	5	8	10
	<i>GM</i>	0,00	13,12 ± 9,09	35,36 ± 4,21	49,77 ± 6,40	<b>52,16 ± 6,59</b>
thyroid	<i>R</i>	3	4	5	8	10
	<i>GM</i>	<b>44,19</b>	42,34 ± 6,13	37,70 ± 7,46	32,98 ± 16,65	36,99 ± 9,57
<b>Средн.</b>	<i>GM</i>	55,73	58,76 ± 3,64	62,01 ± 4,18	64,68 ± 6,02	<b>66,30 ± 4,58</b>



# Лучшие результаты среди 5 запусков

17

Набор данных		ЭК	ЭК + SFLA <sub>add</sub>				Прирост
penbased	<i>R</i>	10	11	12	15	17	
	<i>GM</i>	58,35	61,45	63,19	67,08	<b>69,99</b>	+ 11,64
newthyroid	<i>R</i>	3	4	5	8	10	
	<i>GM</i>	90,20	94,75	95,19	<b>96,24</b>	93,72	+ 6,04
dermatology	<i>R</i>	6	7	8	11	13	
	<i>GM</i>	85,92	87,06	89,15	87,48	<b>89,29</b>	+ 3,37
balance	<i>R</i>	3	4	5	8	10	
	<i>GM</i>	0,00	28,94	42,97	58,78	<b>63,00</b>	+63,00
thyroid	<i>R</i>	3	4	5	8	10	
	<i>GM</i>	44,19	48,58	48,23	<b>52,92</b>	49,58	+8,73

Сравниваемые алгоритмы	p-value	STS	Нулевая гипотеза
GNB - ЭК+SFLA <sub>add</sub>	0,138	-1,483	принять
LRC - ЭК+SFLA <sub>add</sub>	0,345	-0,994	принять
DTC - ЭК+SFLA <sub>add</sub>	0,893	0,135	принять
MLPC - ЭК+SFLA <sub>add</sub>	0,138	1,483	принять
LSVC - ЭК+SFLA <sub>add</sub>	0,5	-0,674	принять
3NN - ЭК+SFLA <sub>add</sub>	0,5	-0,674	принять
ABC - ЭК+SFLA <sub>add</sub>	0,138	-1,483	принять
RFC - ЭК+SFLA <sub>add</sub>	0,893	0,135	принять
GB - ЭК+SFLA <sub>add</sub>	0,893	0,135	принять

- Gaussian naive Bayes (GNB),
- Logistic Regression classifier (LR),
- Decision Tree classifier (DT),
- Multi-Layer Perceptron classifier (MLP),
- Linear Support Vector classifier (LSV),
- K-Nearest Neighbors classifier с  $k=3$  (3NN),
- AdaBoost classifier (AB),
- Random Forest classifier (RF)
- Gradient boosting (GB)

**1 Разработанный алгоритм формирования базы нечетких правил для несбалансированных данных позволил создать базы правил, демонстрирующие большую среднюю геометрическую точность по сравнению с алгоритмом экстремальных значений признаков классов.**

**Средняя геометрическая точность в сформированных классификаторах увеличилась на исследованных бинарных наборах данных на 7,95-12,06% и на многоклассовых наборах на 3,03-10,57%.**

Соответствует пункту 5 паспорта специальности: Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечения.

**2 Разработанный метод нечеткой классификации с настройкой непрерывного вектора весов признаков позволяет улучшить качество классификации.**

**Отбор признаков** – выбор из множества входных признаков релевантного подмножества для уменьшения сложности модели и увеличения точности классификации.

## Методы отбора признаков

### 1. Методы обертки

**Отбор подмножества признаков на основе точности классификации.**

- + получение результатов, адекватных выбранному алгоритму классификации;
- + более высокие результаты по сравнению с методами фильтрации;
- высокая вычислительная сложность;
- риск переобучения.

### 2. Методы фильтрации

**Основаны на анализе свойств и зависимостей во входных данных.**

- + независимость от алгоритма классификации;
- + невысокая вычислительная сложность.
- невысокие результаты по сравнению с методами-обертками.

### 3. Встроенные методы

Отбор признаков, включенный в процесс обучения модели.

### 4. Гибридные методы

# Предлагаемый метод

## Бинарный отбор признаков

0 – признак не участвует в классификаторе;  
1 – участвует в классификаторе

	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$	$s_7$	...
$S_1$	0	1	1	1	0	0	1	...
$S_2$	0	0	0	1	0	1	1	...

Для оптимизации применяются алгоритм «прыгающих лягушек» (SFLA) и «гравитационный поиск» (GSA) со случайной генерацией первоначальной популяции.

В дальнейшем планируется использовать для первоначальной генерации векторов с весами признаков ранжирующий алгоритм фильтрации.

## Непрерывный отбор признаков

$w_j$  – вес  $j$ -ого признака,  $w \in [0; 1]$

	$w_1$	$w_2$	$w_3$	$w_4$	$w_5$	$w_6$	$w_7$	...
$W_1$	0	0,4	1	0,8	0,8	0,2	0,1	...
$W_2$	0,2	0	0	0,9	0	0,1	1	...

Степень принадлежности  $i$ -ому правилу

$$\beta_i = \prod_{j=1}^n (1 - w_j(1 - \mu_j))$$

# Эксперимент

Критерий: общая точность

23

Данные	Признаки	Экземпляры	Классы	ЭК	SFLA	GSA
magic	10	19020	2	71,90	71,21	71,74
wine	13	178	3	88,10	89,35	92,71
heart	13	270	2	71,48	68,15	69,63
cleveland	13	297	5	45,18	50,20	48,56
vowel	13	990	11	43,84	39,60	46,97
penbased	16	10992	10	56,70	55,94	60,99
vehicle	18	846	4	30,39	32,63	44,58
hepatitis	19	80	2	83,42	85,25	80,02
segment	19	2310	7	80,48	74,29	81,73
ring	20	7400	2	64,39	58,41	64,34
twonorm	20	7400	2	96,09	95,96	96,59
thyroid	21	7200	3	7,42	93,82	91,01
wdbc	30	569	2	92,79	93,67	95,08
ionosphere	33	351	2	79,76	85,20	86,05
dermatology	34	358	6	80,74	79,34	88,05
satimage	36	6435	7	60,34	62,21	68,61
spectfheart	44	267	2	78,65	62,95	78,22
spambase	57	4597	2	39,42	63,35	72,42
sonar	60	208	2	62,55	55,26	54,40
optdigits	64	5620	10	13,51	24,61	34,93
coil2000	85	9822	2	94,03	92,85	84,51
Среднее				<b>63,87</b>	<b>68,30</b>	<b>71,96</b>

**2 Разработанный метод нечеткой классификации с настройкой непрерывного вектора весов признаков позволил улучшить качество классификации:**

- при использовании алгоритма «прыгающие лягушки» на 4,43%;**
- при применении «гравитационного алгоритма» на 8,09%.**

Соответствует пункту 5 паспорта специальности: Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечения.



**3 Разработанный гибридный алгоритм настройки параметров нечеткого классификатора позволяет увеличить среднюю геометрическую точность классификации на 3 - 30% при увеличении процента правильной классификации меньшего класса на 3 - 47% по сравнению с традиционным методом построения нечеткого классификатора на исследуемых несбалансированных наборах данных.**

Соответствует пункту 13 паспорта специальности: Применение бионических принципов, методов и моделей в информационных технологиях.

## Алгоритм «гравитационного поиска» (GSA)

Популяция – система частиц, между которыми действует сила притяжения.

Недостатки:

- расчет физических характеристик по громоздким формулам;
- алгоритм представляет собой только глобальный поиск;
- часто «застревает» в локальных экстремумах.

## Алгоритм «прыгающих лягушек» (SFLA)

Популяция – группа лягушек, находящаяся в поиске пищи; популяция разбивается на подгруппы, в которых происходит генерация новых решений.

Недостаток:

- оператор для создания новых векторов не обеспечивает больших изменений в исходном векторе.

$$\mathbf{New} = rand \cdot const \cdot (\mathbf{best} - \mathbf{worst}) + \mathbf{worst}$$

# Объединение в гибрид

## Комбинация

### алгоритмов: GSA

отвечает за глобальный поиск, SFLA выполняет локальный поиск.

## Условия входа в SFLA:

- отсутствие изменений в лучшем решении в течение некоторого количества итераций;
- слишком близкие расстояния между векторами решений;
- заданная итерация.



Рисунок 3 – Схема гибридного алгоритма

Для эксперимента использованы наборы данных из репозитория KEEL (keel.es)

Все наборы имеют по два класса;

- $F_{\text{all}}$  – количество признаков в исходном наборе;
- $F_{\text{use}}$  – количество отобранных признаков бинарным GSA;
- $Ex_{\text{all}}$  – число всех экземпляров;
- $Ex^+$  – число экземпляров меньшего класса,  $Ex^-$  – число экземпляров большего класса;
- $IR$  – коэффициент дисбаланса,  $IR = Ex^- / Ex^+$ .

№	Набор данных	$F_{\text{use}} / F_{\text{all}}$	$Ex_{\text{all}}$	$Ex^+$	$Ex^-$	$IR$
1	vehicle0	11/18	846	199	647	<b>3,25</b>
2	newthyroid2	3/5	215	35	180	<b>5,14</b>
3	segment0	6/19	2308	329	1979	<b>6,02</b>
4	page-blocks0	2/10	5472	559	4913	<b>8,79</b>
5	vowel0	6/13	988	90	898	<b>9,98</b>
6	cleveland-0vs4	2/13	177	13	164	<b>12,62</b>
7	ecoli4	4/7	336	20	316	<b>15,8</b>
8	yeast4	2/8	1484	51	1433	<b>28,1</b>

## Порядок эксперимента:

- структура классификатора формируется алгоритмом на основе экстремумов классов;
- классификатор перестраивается на отобранных  $GSA_b$  признаках;
- запускается непрерывный алгоритм оптимизации параметров.

## Критерии качества классификации:

- $Acc$  – общая точность классификации;
- $GM$  – средняя геометрическая точность;
- $A+$  – точность на экземплярах меньшего класса;
- $A-$  – точность на экземплярах большего класса.

Данные	vehicle0, $IR = 3,25$				newthyroid2, $IR = 5,14$			
Алг.	До оптим.	GSA	SFLA	GSA+SFLA	До оптим.	GSA	SFLA	GSA+SFLA
<i>Acc</i>	77,89	84,00	93,27	93,39	96,28	99,53	100,00	100,00
<i>GM</i>	64,74	80,58	93,47	94,09	95,41	99,32	100,00	100,00
<i>A+</i>	48,72	75,88	93,95	95,50	96,67	99,05	100,00	100,00
<i>A-</i>	86,86	86,50	93,05	92,74	94,29	99,63	100,00	100,00
Данные	segment0, $IR = 6,02$				page-blocks0, $IR = 8,79$			
Алг.	До оптим.	GSA	SFLA	GSA+SFLA	До оптим.	GSA	SFLA	GSA+SFLA
<i>Acc</i>	97,14	99,02	99,35	99,35	48,08	92,85	93,07	93,70
<i>GM</i>	95,77	98,66	99,24	99,49	69,65	81,52	85,16	85,93
<i>A+</i>	93,93	98,18	99,09	99,70	24,85	70,60	76,39	77,28
<i>A-</i>	97,68	99,16	99,39	99,29	98,78	95,39	94,97	95,56

*Acc* – общая точность классификации;  
*GM* – средняя геометрическая точность;

*A+* – точность на экземплярах меньшего класса;  
*A-* – точность на экземплярах большего класса.

# Сравнение результатов – часть 2

31

Данные	vowel0, $IR = 9,98$				cleveland-0_vs_4-5, $IR = 12,62$			
Алг.	До оптим.	GSA	SFLA	GSA+SFLA	До оптим.	GSA	SFLA	GSA+SFLA
<i>Acc</i>	85,62	97,74	97,97	95,95	89,81	93,02	92,62	94,89
<i>GM</i>	84,48	97,41	97,38	97,26	85,97	86,17	78,81	92,03
<i>A+</i>	83,33	97,04	96,67	98,89	83,33	82,05	66,67	90,00
<i>A-</i>	85,85	97,81	98,11	95,66	90,27	93,90	94,53	95,15
Данные	ecoli4, $IR = 15,8$				yeast4, $IR = 28,1$			
Алг.	До оптим.	GSA	SFLA	GSA+SFLA	До оптим.	GSA	SFLA	GSA+SFLA
<i>Acc</i>	96,14	97,92	98,21	94,63	86,39	91,19	90,43	90,36
<i>GM</i>	84,54	93,88	94,04	97,06	71,39	80,40	79,97	80,98
<i>A+</i>	75,00	90,00	90,00	100,00	59,09	70,59	70,36	72,36
<i>A-</i>	97,48	98,42	98,73	94,29	87,37	91,93	91,14	91,00

*Acc* – общая точность классификации;  
*GM* – средняя геометрическая точность;

*A+* – точность на экземплярах меньшего класса;  
*A-* – точность на экземплярах большего класса.

1 Разработанный гибридный алгоритм настройки параметров нечеткого классификатора позволил **увеличить среднюю геометрическую точность классификации на 3 - 30% при увеличении процента правильной классификации меньшего класса на 3 - 47%** по сравнению с традиционным методом построения на исследованных несбалансированных наборах данных.

Соответствует пункту 13 паспорта специальности: Применение бионических принципов, методов и моделей в информационных технологиях.



**1 Разработан новый алгоритм дополнения базы нечетких правил на основе метаэвристики «прыгающие лягушки».**

**2 Разработан новый метод нечеткой классификации с настройкой непрерывного вектора весов признаков.**

**3 Разработан гибридный алгоритм настройки параметров нечетких классификаторов несбалансированных данных, отличающийся совмещением алгоритма «прыгающих лягушек» и «гравитационного поиска».**

## **9 публикаций, индексируемых в Scopus:**

1. Bardamova, M.B., Hodashinsky, I.A. Optimization of fuzzy classifier parameters with a combination of gravitational search algorithm and shuffled frog leaping algorithm // Journal of Physics: Conference Series, 2020, 1611, 012068
2. Hodashinsky I. A., Bardamova M. B., Kovalev V.S. Using Shuffled Frog-Leaping Algorithm for Feature Selection and Fuzzy Classifier Design // Scientific and Technical Information Processing, 2019, Vol. 46, No. 6, pp. 381–387.
3. Bardamova M.; Konev A.; Hodashinsky I.; Shelupanov A. Gravitational search for designing fuzzy rule-based classifiers for handwritten signature verification // Journal of Communications Software and Systems. – 2019. – 15(3). – С. 254-261;
4. Bardamova M.; Konev A.; Hodashinsky I.; Shelupanov A. Application of the Gravitational Search Algorithm for Constructing Fuzzy Classifiers of Imbalanced Data // Symmetry. – 2019. – 11. – С. 1458;
5. и т.д.

## **3 публикации, индексируемые в ВАК:**

1. Ходашинский, И.А. Аутентификация пользователя по динамике подписи на основе нечеткого классификатора / И.А. Ходашинский, Е.Ю. Костюченко, К.С. Сарин, А.Е. Анфилофьев, М.Б. Бардамова, С.С. Самсонов, И.В. Филимоненко // Компьютерная оптика. – 2018. – Т. 42, №4. – С. 657-666.
2. Ходашинский И. А., Бардамова М.Б., Ковалев В.С. Отбор признаков и построение нечеткого классификатора на основе алгоритма прыгающих лягушек // Искусственный интеллект и принятие решений. – 2018. – № 1. – С. 76-84.
3. Ходашинский И. А. Построение нечеткого классификатора алгоритмом гравитационного поиска / И. А. Ходашинский, М. Б. Бардамова, В. С. Ковалев // Доклады ТУСУР. – 2017. – Т. 20, № 2. – С. 84–87.

26 публикаций, индексируемых в РИНЦ

## Выступления на конференциях:

1. Выступление с докладом «Construction of fuzzy classifiers with shuffle frog leaping algorithm for handwritten signature authentication» на Всероссийской научно-технической конференции «Научная сессия ТУСУР», Томск, 22-24 мая 2019 г.
2. Выступление с докладом «Построение нечетких классификаторов рукописных подписей с применением метаэвристических алгоритмов» на XVI Международной конференции студентов, аспирантов и молодых ученых "Перспективы развития Фундаментальных наук", Томск, 23-26 апреля 2019 г.
3. Выступление с докладом «Constructing the fuzzy classifier for analyzing imbalanced network traffic data with the gravitational search algorithm2 // на XV Международной научно-практической конференции «Электронные средства и системы управления», Томск, 20-22 ноября 2019 г.
4. Выступление с докладом «Construction of fuzzy classifiers with Gravitational Search algorithm for handwritten signature authentication» на IV Всероссийском Форуме "Наука будущего - наука молодых", Сочи, 14-17 мая 2019 г.
5. Выступление с докладом «Построение нечеткого классификатора на основе гравитационного алгоритма для аутентификации пользователя по динамике подписи» на XVIII Всероссийском конкурсе-конференции студентов и аспирантов по Информационной безопасности «SIBINFO-2018», Томск, 19 апреля 2018 г.

И другие 18 конференций в Томске, Санкт-Петербурге, Москве, Екатеринбурге, Нижнем Новгороде.

## Учебно-методические материалы:

Теория информации: Методические указания для выполнения практических и самостоятельных работ / Ходашинский И. А., Бардамова М. Б. — 2018. 64 с.

## Преподавательская деятельность:

Суммарно около 1188+ часов:

- «Теория информации» – весна 2017, 2018, 2019 гг. – 387 часов;
- «Безопасность систем баз данных» - весна 2020 г. – 132 часа;
- «Информатика» – осень 2018, 2019, 2020 гг. – 669+ часов.

## Стипендии:

1. Стипендия Президента Российской Федерации для молодых ученых и аспирантов, осуществляющих перспективные научные исследования и разработки по приоритетным направлениям модернизации российской экономики на 2019-2021 гг.
2. Стипендия Правительства Российской Федерации на 2018/19, 2019/20, 2020/21 год.

## Участие в грантах:

1. Научный проект при поддержке РФФИ «Методы построения нечетких классификаторов несбалансированных данных на основе алгоритма гравитационного поиска» (№19-37-90064, код «Аспиранты»), 2019-2021 гг.
2. Научный проект при поддержке РФФИ «Методы и инструментальные средства построения самообучающихся систем, основанных на нечетких правилах» (№16-07-00034 А), 2016-2018 гг.
3. Государственное задание ТУСУР на 2017–2019 гг. Проект № 2.8172.2017/8.9 «Методы и модели определения уровня защищенности информационных систем»
4. Государственное задание Министерства образования и науки Российской Федерации на 2017-2019 гг. Проект № 8.9628.2017/8.9 «Теоретические основы человеко-машинных интерфейсов»

## Свидетельства о регистрации ПО для ЭВМ:

1. Программа отбора признаков для нечеткого классификатора на основе бинарного алгоритма гравитационного поиска со статическими функциями трансформации. Бардамова М.Б., Ходашинский И.А. Свидетельство о государственной регистрации программы для ЭВМ №: 2019612574. Дата государственной регистрации в реестре Роспатента ПрЭВМ: 22.02.2019. Заявка № 2019610957. Дата приоритета: 05.02.2019. Правообладатель: ТУСУР.
2. Программа настройки параметров нечеткого классификатора на основе гравитационного поиска. Бардамова М.Б., Ходашинский И.А. Свидетельство о государственной регистрации программы для ЭВМ № 2018614316. Дата государственной регистрации в реестре Роспатента ПрЭВМ: 04.04.2018 г. Заявка № 2017662551. Дата приоритета: 04.12.2017. Правообладатель: ТУСУР.

Спасибо за внимание.