

УДК 004.85

М.О. Светлаков, И.Г. Боровской

## Отбор признаков в нечетких системах бинарной классификации типа Ангелова–Ягера при обработке потоковых данных

Важной проблемой в сфере машинного обучения является поддержание обученных моделей в актуальном состоянии на основе потоковых данных, так как существующие решения не всегда способны обновляться и обрабатывать данные в инкрементном режиме. Одно из существующих решений с поддержкой инкрементного обучения – нечеткие системы бинарной классификации типа Ангелова–Ягера первого порядка. Недостатком данных систем является то, что в режиме вывода-предсказания система работает на полном наборе признаков, даже если не все признаки являются информативными. В настоящей работе для указанных нечетких систем предложен комплексный способ расчета важности признаков и проведен вычислительный эксперимент отбора признаков при обработке потоковых данных на наборах данных по тематике обнаружения спама, фишинговых сайтов и атак на сетевые соединения, выявлена статистически значимая разница по точности и числу правил в пользу применения предложенного способа расчета важности признаков.

**Ключевые слова:** нечеткие системы типа Ангелова–Ягера, отбор признаков, потоковые данные, бинарная классификация.

**DOI:** 10.21293/1818-0442-2025-28-4-50-56

Для решения задачи бинарной классификации используются различные инструменты, такие как методы, основанные на правилах, деревья решений, нейронные сети, ансамблевые методы. Многие из этих методов были разработаны для построения моделей классификации на основе пакетной обработки данных, где возможно несколько обращений к сохраненным данным, которые невозможно хранить в полном виде в случае потоков данных. Также использование статических моделей для решения задач классификации и кластеризации потоковых данных может привести к деградации точности с течением времени, поскольку моделируемая моделью зависимость может меняться [1–3].

В таком случае необходимо выполнение оценки актуальности прошлых данных и обновление моделей при появлении новых – метод инкрементного обучения. Согласно этому методу, прогностическая модель может дообучиться с приходом нового экземпляра данных без необходимости полного переобучения как со старой, так и с новой информацией [4]. Одно из существующих решений с поддержкой инкрементного обучения – нечеткие системы бинарной классификации типа Ангелова–Ягера (Autonomous Learning Multi-Model Systems), которые были представлены в [5, 6].

Данные системы основаны на методологии эмпирического анализа данных [7] и состоят из облаков правил (нечетких правил) с фокальными точками, которые непрерывно создаются, обновляются или удаляются в процессе обработки потоков данных для учета проблемы дрейфа понятий.

Указанные системы подразделяются на два вида в зависимости от содержания консеквентной части и механизма вывода: 1) Autonomous Learning Multi-Model Classifier of 0-Order (ALMMo-0); 2) Autonomous Learning Multi-Model Classifier of First-Order (ALMMo-1). В данной работе рассматривается только модель первого порядка – ALMMo-1, являю-

щаяся более точной [6]. Недостатком данной системы является то, что в режиме вывода-предсказания система работает на полном наборе признаков, даже если не все признаки являются информативными. В случае потоковых данных набор информативных признаков также может меняться с течением времени. Немногие существующие методы отбора признаков поддерживают потоковые данные, среди существующих методов из наименьших по вычислительной затратности можно выделить метод OFFESEL из [8].

В [6] указана, но не исследована возможность отбора признаков для ALMMo-1 на основе особенностей системы, а именно весов в консеквентной части правил. Использование таких структурных особенностей системы в совокупности с OFFESEL представляется перспективным ввиду довольно низких вычислительных затрат.

Цель работы заключалась в реализации отбора признаков в нечетких системах бинарной классификации типа Ангелова–Ягера при обработке потоковых данных.

Научный вклад работы представлен следующими результатами:

1) предложен комплексный способ расчета важности признаков для нечеткой системы бинарной классификации типа Ангелова–Ягера первого порядка, учитывающий изменения потоковых данных и основанный на параметрических значениях указанной нечеткой системы и бисериальном коэффициенте корреляции;

2) проведен вычислительный эксперимент предложенного способа расчета важности признаков при обработке потоковых данных на наборах данных по тематике обнаружения спама, фишинговых сайтов и атак на сетевые соединения, выявлена статистически значимая разница по точности и числу правил в пользу применения предложенного способа расчета важности признаков.

**Описание нечеткой системы ALMMo-1**

Далее приведено описание важных для отбора признаков принципов работы ALMMo-1. ALMMo-1 состоит из облаков-правил вида

$$R_j : \text{если } (\mathbf{x}_K \sim \mathbf{L}_{j,K}) \text{ то } (y_{j,K} = \mathbf{u}_K^T \mathbf{a}_{j,K}), \quad (1)$$

где  $\mathbf{x}_K$  – входной вектор значений признаков в момент времени  $K$  ( $\mathbf{x}_K = (x_{1,K}, \dots, x_{M,K})$ ,  $K = 1$  в начале отсчета),  $M$  – число признаков,  $\mathbf{L}_{j,K}$  – фокальная точка  $j$ -го облака ( $\mathbf{L}_{j,K} = [l_{j,K,1}, \dots, l_{j,K,M}]$ ),  $y_{j,K}$  – вывод  $j$ -го облака для  $\mathbf{x}_K$ ,  $\mathbf{u}_K^T = [1, x_{1,K}, \dots, x_{M,K}]$ ,  $\mathbf{a}_{j,K} = [a_{j,K,0}, a_{j,K,1}, \dots, a_{j,K,M}]$ .

Общий вывод системы  $y_K^{\text{out}}$  для  $\mathbf{x}_K$  вычислялся следующим образом для задачи бинарной классификации:

$$y_K^{\text{out}} = \begin{cases} 1, & \text{если } v_K < 0,5; \\ 0, & \text{иначе;} \end{cases} \quad (2)$$

$$v_K = \sum_{j=1}^{N_K} \lambda_{j,K} \mathbf{u}_K^T \mathbf{a}_{j,K} = \Psi_K^T \mathbf{A}_K, \quad (3)$$

где  $N_K$  – число облаков в момент времени  $K$ ,  $\lambda_j$  – степень активации  $j$ -го облака,  $\Psi_K^T = [\lambda_{1,K} \mathbf{u}_K^T, \lambda_{2,K} \mathbf{u}_K^T, \dots, \lambda_{N,K} \mathbf{u}_K^T]$ .

В алгоритме выделены понятия *локальных* фокальных точек  $\mathbf{L}_j$  – центров облаков, а также одной *глобальной* фокальной точки  $\mathbf{G}_j$ .  $\mathbf{L}_j$  рекуррентно обновлялся следующим образом:

$$\mathbf{L}_{K,j} = \frac{S_{K-1}^j}{S_{K-1}^j + 1} \mathbf{L}_{K-1,j} + \frac{1}{S_{K-1}^j + 1} \mathbf{x}_K, \quad (4)$$

$$\mathbf{L}_{1,j} = \mathbf{x}_j^{\text{init}}, \quad (5)$$

где  $\mathbf{x}_j^{\text{init}}$  – образец данных, которым инициализировано  $j$ -е облако в момент времени  $K_j^{\text{init}}$ ,  $S_K^j$  – число образцов данных, которые были использованы для обновления  $j$ -го облака ( $S_K^j = 1$ ). Глобальная фокальная точка обновлялась следующим образом:

$$\mathbf{G}_K = \frac{K-1}{K} \mathbf{G}_{K-1} + \frac{1}{K} \mathbf{x}_K, \quad (6)$$

$$\mathbf{G}_1 = \mathbf{x}_1. \quad (7)$$

Консеквентная часть обновлялась следующим образом:

$$\mathbf{C}_{j,K+1} = \mathbf{C}_{j,K} - \frac{\lambda_{j,K} \mathbf{C}_{j,K} \mathbf{u}_K \mathbf{u}_K^T \mathbf{C}_{j,K}}{1 + \lambda_{j,K+1} \mathbf{u}_K \mathbf{C}_{j,K} \mathbf{u}_K^T}, \quad (8)$$

$$\mathbf{d}_{j,K+1} = \lambda_{j,K} \mathbf{C}_{j,K+1} \mathbf{u}_K (y_K^{\text{true}} - \mathbf{u}_K^T \mathbf{a}_{j,K}), \quad (9)$$

$$\mathbf{a}_{j,K+1} = \mathbf{a}_{j,K} + \mathbf{d}_{j,K+1}, \quad (10)$$

где  $y_K^{\text{true}}$  – истинная классификационная метка образца  $\mathbf{x}_K$ .

**Отбор признаков**

Основная цель ранжирования и отбора признаков – удалить нерелевантные и избыточные признаки

из моделей машинного обучения, оставив только наиболее информативные. Важность признаков в динамических потоках данных может меняться со временем. Недостатком системы ALMMo-1 является то, что в режиме вывода-предсказания система работает на полном наборе признаков, даже если не все признаки являются информативными.

В [6] указана, но не исследована возможность отбора признаков для ALMMo-1 на основе особенностей системы, а именно весов в консеквентной части правил, на основе которых важность признаков  $w$  обновляется следующим образом:

$$w_{i,K}^{\text{original}} = \left( \sum_{j=1}^{N_K} |\mathbf{a}_{i,j,K}| \right) / \left( \sum_{i=1}^M \sum_{j=1}^{N_K} |\mathbf{a}_{i,j,K}| \right), \quad (11)$$

где  $w_{i,K}$  – важность  $i$ -го признака в интервале  $[0, 1]$ . Обозначим данный способ формирования важности признаков как *original*. Если  $w_{i,K}$  меньше фиксированного значения или если  $i$ -й признак не входит в  $k$  наибольших по важности признаков, соответствующие признаку столбцы и строки удаляются из консеквентов  $\mathbf{C}_{i,K}$  и  $\mathbf{a}_{i,K}$  всех правил. так как в консеквентной части фактически содержится регрессионная модель, удаление из неё большого числа регрессоров без корректировки остальных коэффициентов (например, если есть жесткое ограничение по максимальному числу признаков) может привести к значительной деградации точности, поэтому в настоящей работе предложено запоминать последний пришедший образец ( $\mathbf{x}_k, y_k$ ) и после каждого момента формирования классификатора с усеченным набором признаков на основе *original* дополнительно обновлять веса консеквентной части с помощью (8)–(10).

Немногие существующие методы отбора признаков поддерживают потоковые данные, среди существующих методов из наименьших по вычислительной затратности можно выделить метод OFFESEL (OF) из [8]. Авторы хранили небольшой набор последних данных («окно данных») и на основе минимальных и максимальных значений признаков в этом наборе нормализовали каждый числовой признак в диапазон  $[0, 1]$ . После нормализации авторы вычисляли среднее значение нормализованного  $i$ -го признака  $\mu_i$  и определяли важность признака следующим образом:

$$w_{i,K}^{\text{OF}} = (|\mu_{i,K} - 0,5|) / \left( \sum_{i=1}^M |\mu_{i,K} - 0,5| \right). \quad (12)$$

В настоящей работе предложено совместить предложенные два подхода выше с другими, чтобы сформировать комплексный способ расчета важности признаков, который обозначим как *mixed*. Для его реализации предложено дополнительно отслеживать стандартное отклонение (*STD*) на основе следующих формул:

$$STD_{K,i} = \frac{(K-1)STD_{K-1,i} + (x_{K,i} - STD_{K-1,i})^2}{K}, \quad (13)$$

$$STD_{1,i} = 0, \quad (14)$$

$$w_{i,K}^{std} = (STD_{i,K}) / (\sum_{i=1}^M STD_{i,K}). \quad (15)$$

Метод OFFESEL адаптирован для ALMMo-1 следующим образом:

$$w_{i,K}^{OF} = (|G_{i,K} - 0,5|) / (\sum_{i=1}^M |G_{i,K} - 0,5|). \quad (16)$$

Также в рамках адаптации метода OFFESEL реализована потоковая нормализация данных из [9], которая в отличие от нормализации в [8] работает не на «окне данных», а обновляет минимумы, максимумы и параметры системы на каждом входном образце.

Также введены следующие дополнительные компоненты:

$$\bar{l}_{i,K} = (\sum_{i=1}^{N_K} l_{i,j,K}) / N_K, \quad (17)$$

$$STD_{i,K}^{centre} = \sqrt{(\sum_{i=1}^{N_K} (l_{i,j,K} - \bar{l}_{i,K})^2) / (N_K - 1)}, \quad (18)$$

$$w_{i,K}^{centre} = (STD_{i,K}^{centre}) / (\sum_{i=1}^M STD_{i,K}^{centre}), \quad (19)$$

$$\beta_{i,K} = \frac{(G_{i,K}^{pos} - G_{i,K}^{neg})}{STD_{i,K}} \sqrt{\frac{n^{pos} n^{neg}}{K}}, \quad (20)$$

$$w_{i,K}^{corr} = \beta_{i,K} / (\sum_{i=1}^M \beta_{i,K}), \quad (21)$$

где  $G_{i,K}^{pos}$  и  $G_{i,K}^{neg}$  – средние значения  $i$ -го признака среди обработанных образцов 1-го класса ( $y_K = 1$ ) и 2-го класса ( $y_K = 0$ ) соответственно, обновляемые аналогично (7), (8);  $n^{pos}$  и  $n^{neg}$  – число обработанных образцов 1-го и 2-го класса соответственно.  $\beta_{i,K}$  представляет собой адаптированный точечный бисериальный коэффициент корреляции. Далее приведена итоговая формула расчета весов для комплексного способа расчета важности признаков mixed:

$$w_{i,K}^{new} = \frac{(w_{i,K}^{OF} + w_{i,K}^{centre} + w_{i,K}^{corr} + w_{i,K}^{original} + w_{i,K}^{std})}{5}, \quad (22)$$

$$w_{i,K}^{mixed} = ((K-1)w_{i,K-1}^{mixed} + w_{i,K}^{new}) / K, \quad (23)$$

$$w_{i,1}^{mixed} = w_{i,1}^{new}. \quad (24)$$

### Вычислительный эксперимент

Для построения моделей машинного обучения на основе потоковых данных были выбраны 3 набора данных по тематике обнаружения спама, фишинговых сайтов и атак на сетевые соединения; обозначения наборов данных и их характеристики приведены в табл. 1. Все 3 набора данных содержат только два класса.

Наборы данных были разделены на обучающие и тестовые данные согласно схеме 10-кратной перекрестной проверки [13]. Для того чтобы оценить

эффективность моделей обработки потоковых данных, необходимо проводить оценку критериев эффективности на протяжении всего цикла обучения [14, 15]. Для этого определяют номера образцов потока, с появлением которых будут проводиться оценки. Обозначим множество таких номеров  $T = [t_1, t_2, \dots, t_m]$ . В настоящей работе оценки проводились каждые  $m$  обработанных образцов,  $m = 30$ . При получении образца  $t_i$  выполняется оценка эффективности текущей модели на всей тестовой выборке.

Таблица 1

Характеристики наборов данных

Набор данных	Число образцов	Число признаков	Процентное соотношение классов
UNSW-NB15 [10]	257673	43	36/64
Phishing Websites [11]	88647	112	65/35
Spambase [12]	4597	57	60/40

Таким образом, после окончания эксперимента на одном разбиении кросс-валидации зафиксировано 30 оценок критерия [ $perf_1, perf_2, \dots, perf_{30}$ ]. На каждом разбиении данных на обучающую и тестовую выборки процедура повторялась, итоговые оценки критериев  $perf_i$  определены средним значением по 10 разбиениям перекрестной проверки. В качестве критериев эффективности использовались F-мера (F1) [13] и число отобранных признаков.

Отбор признаков в рамках эксперимента был реализован следующим образом. Модель обучалась на полном наборе признаков, так как признак, на каком-либо этапе признанный неинформативным, в дальнейшем может снова стать информативным. При получении образца  $t_i$ , когда выполняется оценка эффективности текущей модели на всей тестовой выборке, создавалась полная копия обученной к этому моменту времени модели.

Именно для этой копии проводился отбор признаков. Для этого хранились последние 300 образцов данных (по 150 на класс) из потока и по одному удалялись признаки с наименьшим значением важности  $w$  (рассчитываемого по (11) или (23)). Изначально  $z_i^{new} = M$ . После каждого удаления признака  $z_i^{new}$  уменьшался на единицу, обновлялись веса консеквентной части модели-копии с помощью последнего запомненного образца входных данных и (8)–(10), для модели с  $z_i^{new}$  лучших по значению  $w$  признаков выполнялась оценка точности на 300 запомненных последних образцах. Если новое значение F1 было хуже более чем на 0,01 по сравнению со значением F1 на  $z_i^{new} + 1$  признаков, тогда запоминалось предыдущее значение  $z_i^{new}$ . Итоговая оценка  $perf_i$  рассчитывалась на всей тестовой выборке с моделью-копией, где использовались  $\text{floor}(z_i^{mem})$  лучших по значению  $w$  признаков,  $z_i^{mem}$  обновлялось следующим образом:

$$z_i^{mem} = ((i - 1)z_{i-1}^{mem} + z_i^{new}) / i, i \in [1, 30], \quad (25)$$

$$z_1^{mem} = z_1^{new}. \quad (26)$$

Результаты оценок эффективности приведены на рис. 1 и 2. Средние значения и среднеквадратическое отклонение (СКО) оценок приведены в табл. 2.

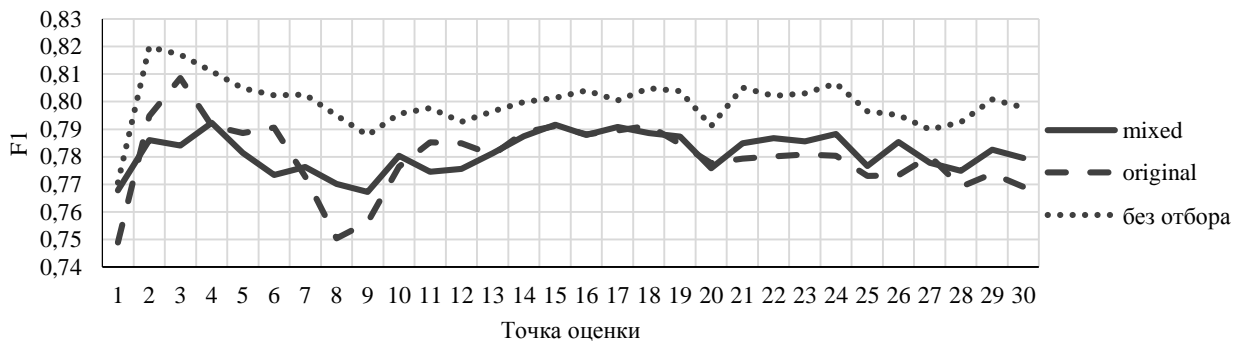
**Анализ результатов**

Для набора данных Spambase предложенный способ расчета важности признаков mixed работает лучше, так как отобрал меньше признаков (статистически значимая разница при сравнении по числу признаков с использованием парного статистического теста Уилкоксона,  $p\text{-value} < 0,001$ ), чем original, при этом по точности они не различимы (при сравнении по F1,  $p\text{-value} = 0,221$ ).

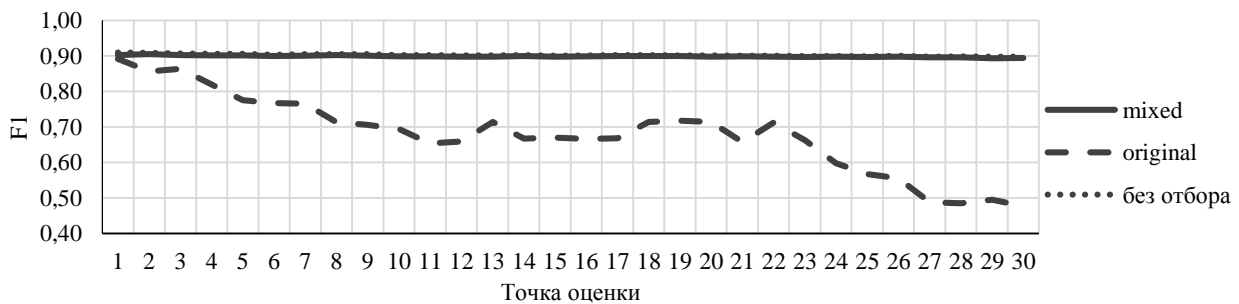
Для набора данных UNSW-NB15 original отобрал меньше признаков, однако при этом точность сильно деградировала (см. рис. 1, б), а при использовании mixed точность ухудшилась лишь на 0,0033 (статистически значимая разница при сравнении по F1,  $p\text{-value} < 0,001$ ) в среднем по сравнению с построением без отбора признаков.

Для набора данных Phishing Websites при использовании original точность также деградировала (см. рис. 1, в). На этом наборе данных mixed превзошел analog original как по точности (статистически значимая разница при сравнении по F1,  $p\text{-value} < 0,001$ ), так и отобрал меньше признаков (статистически значимая разница при сравнении по числу признаков,  $p\text{-value} < 0,001$ ). Наиболее вероятной причиной подобного падения точности на наборах данных Phishing Websites и UNSW-NB15 предполагается формирование способом original (на основе только весов в консеквентной части правил) такого вектора важности признаков  $w$ , на основе которого часть информативных признаков (имеющих большое влияние на результат классификации на тестовых данных) не отбирается в ходе указанной ранее процедуры выбора признаков на 300 запомненных образцах из обучающего потока данных.

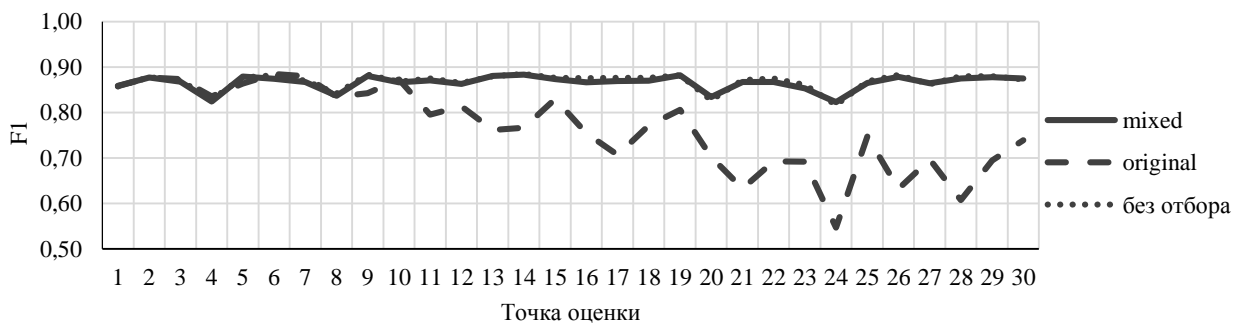
Авторский способ расчета важности признаков mixed таким недостатком на указанных наборах данных не обладает.



а



б



в

Рис. 1. Значения F1 на протяжении процедуры оценки эффективности для набора данных: Spambase – а; UNSW-NB15 – б; Phishing Websites – в

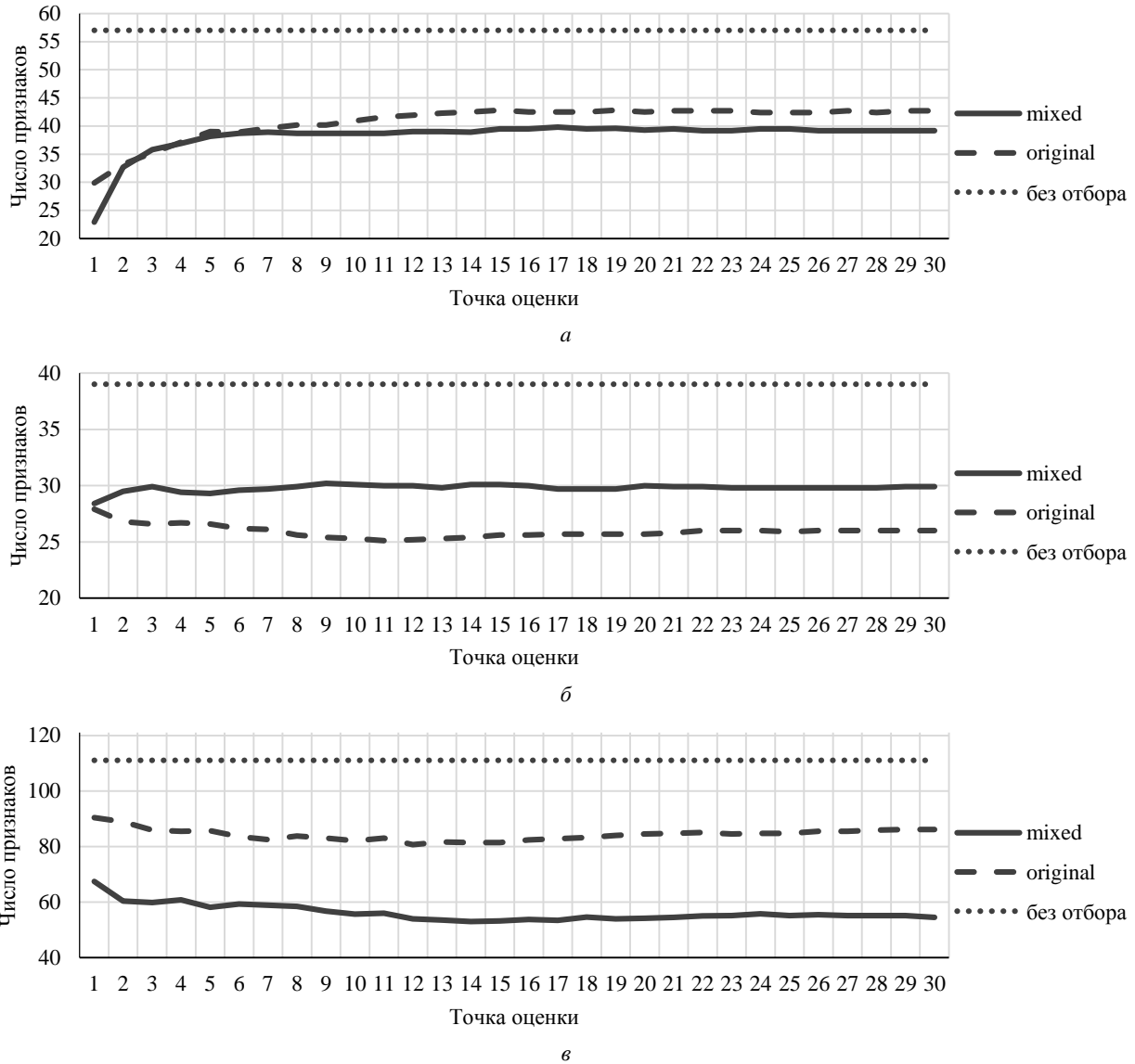


Рис. 2. Число признаков на протяжении процедуры оценки эффективности для набора данных: Spambase – а; UNSW-NB15 – б; Phishing Websites – в

Таблица 2

Средние значения и СКО оценок эффективности				
Способ рас- чета важности признаков	Наборы данных			
	F1		Число признаков	
	среднее	СКО	среднее	СКО
Spambase				
Mixed	0,7814	0,0069	38,1967	3,1526
Original	0,7800	0,0126	40,7967	3,1042
Без отбора	0,7996	0,0089	57,0000	0,0000
Unsw-nb15				
Mixed	0,8989	0,0025	29,7833	0,3257
Original	0,6796	0,1079	25,9300	0,5599
Без отбора	0,9023	0,0030	39,0000	0,0000
Phishing websites				
Mixed	0,8658	0,0160	56,2433	3,0271
Original	0,7672	0,0901	84,3333	2,1050
Без отбора	0,8683	0,0172	111,0000	0,0000

Таким образом, предложенный способ расчета важности признаков mixed позволил без больших потерь точности лучше отбирать признаки на всех наборах данных по сравнению с аналогом original.

**Заключение**

Недостатком инкрементно обучающихся нечетких систем бинарной классификации типа Ангелова–Ягера первого порядка является то, что в режиме вывода система работает на полном наборе признаков, даже если не все признаки являются информативными. В настоящей работе для данных систем предложен комплексный способ расчета важности признаков и проведен вычислительный эксперимент по отбору признаков в потоковом режиме на трех наборах данных, выявлена статистически значимая разница по точности и числу правил в пользу применения предложенного способа расчета важности признаков.

Исследование выполнено за счет гранта Российского научного фонда № 24-21-00168, <https://rscf.ru/project/24-21-00168/>.

### Литература

1. Evolving anomaly detection for network streaming data / W. Xiaolan, M.M. Ahmed, M.N. Husex, Z. Qian // *Information Sciences*. – 2022. – Vol. 608. – P. 757–777.
2. Anomaly detection in streaming data: A comparison and evaluation study / F.I. Vázquez, A. Hartl, T. Zseby, A. Zimek // *Expert Systems with Applications*. – 2023. – Vol. 233. – P. 120994.
3. Cluster based active learning for classification of evolving streams / D. Himaja, V. Dondeti, S. Uppalapati, S. Virupaksha // *Evol. Intel.* – 2023. – Vol. 17, No. 4. – P. 2167–2191.
4. Orlov A.A. Development and research of a neural network alternate incremental learning algorithm / A.A. Orlov, E.S. Abramova // *Computer Optics*. – 2023. – Vol. 47, No. 3. – P. 491–498.
5. Angelov P. Autonomous learning multi-model classifier of 0-Order (ALMMo-0) / P. Angelov, X. Gu // 2017 *Evolving and Adaptive Intelligent Systems (EAIS)*. – 2017. – P. 1–7.
6. Angelov P. Autonomous Learning Multimodel Systems from Data Streams / P. Angelov, X. Gu, J.C. Principe // *IEEE Trans. Fuzzy Syst.* – 2018. – Vol. 26, No. 4. – P. 2213–2224.
7. Angelov P. Empirical Data Analytics / P. Angelov, X. Gu, D. Kangin // *Int. J. Intell. Syst.* – 2017. – Vol. 32, No. 12. – P. 1261–1284.
8. Hochma Y. Fast online feature selection in streaming data / Y. Hochma, M. Last // *Mach Learn.* – 2025. – Vol. 114, No. 1. – P. 1–35.
9. Angelov P. *Autonomous Learning Systems*. – Hoboken: John Wiley and Sons, 2001. – 224 p.
10. Moustafa N. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set) / N. Moustafa, J. Slay // 2015 *Military Communications and Information Systems Conference (MilCIS)*. – Canberra, Australia, 2015. – P. 1–6.
11. Vrbaničič G. Datasets for phishing websites detection / G. Vrbaničič, I.Jr. Fister, V. Podgorelec // *Data in Brief*. – 2020. – Vol. 33. – P. 106438.
12. Spambase [Электронный ресурс]. – Режим доступа: <https://archive.ics.uci.edu/dataset/94/spambase> (дата обращения: 02.10.2025).
13. *Statistical Learning* / G. James, D. Witten, T. Hastie, R. Tibshirani // *Springer Texts in Statistics*. – 2021. – P. 15–57.
14. Data stream classification with novel class detection: a review, comparison and challenges / S.U. Din, J. Shao, J. Kumar, C.B. Mawuli // *Knowl Inf Syst.* – 2021. – Vol. 63, No. 9. – P. 2231–2276.
15. Losing V. Incremental on-line learning: A review and comparison of state of the art algorithms / V. Losing, B. Hammer, H. Wersing // *Neurocomputing*. – 2018. – Vol. 275. – P. 1261–1274.

### Светлаков Михаил Олегович

Канд. техн. наук, доцент каф. компьютерных систем в управлении и проектировании (КСУП) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР) Ленина пр-т, 40, г. Томск, Россия, 634050  
ORCID: 0000-0002-5060-0958  
Тел.: +7 (382-2) 41-47-17  
Эл. почта: svetlakov.m4@gmail.com

### Боровской Игорь Георгиевич

Д-р физ.-мат. наук, проф., зав. каф. экономической математики, информатики и статистики (ЭМИС) ТУСУРа Ленина пр-т, 40, г. Томск, Россия, 634050  
ORCID: 0000-0003-2707-7203  
Тел.: +7 (382-2) 90-01-85  
Эл. почта: igor.g.borovskoi@tusur.ru

Поступила в редакцию: 11.10.2025.

Принята к публикации: 13.01.2026.

Svetlakov M.O., Borovskoy I.G.

### Feature selection in Angelov-Yager binary classification fuzzy systems in the process of stream data processing

An important problem in machine learning is keeping trained models up to date using data streams, as existing solutions are not always capable of updating and processing data incrementally. One of the existing solutions with incremental learning support is first-order Angelov-Yager-type fuzzy binary classification systems. The disadvantage of this system is that in inference mode the system operates on a full set of features, even if not all features are relevant. This paper proposes a comprehensive method for calculating feature importance for the specified fuzzy system and presents an experimental study results of feature selection for processing data streams on datasets, thematically dedicated to spam detection, phishing sites, and network connection attacks. A statistically significant difference in accuracy and number of rules was found in favor of using the proposed method for calculating feature importance.

**Keywords:** Angelov-Yager fuzzy systems, feature selection, data streams, binary classification.

**DOI:** 10.21293/1818-0442-2025-28-4-50-56

### References

1. Xiaolan W., Ahmed M.M., Husex M.N., Qian Z. Evolving anomaly detection for network streaming data. *Information Sciences*, 2022, vol. 608, pp. 757–777.
2. Vázquez F.I., Hartl A., Zseby T., Zimek A. Anomaly detection in streaming data: A comparison and evaluation study. *Expert Systems with Applications*, 2023, vol. 233, pp. 120994.
3. Himaja D., Dondeti V., Uppalapati S., Virupaksha S. Cluster based active learning for classification of evolving streams. *Evol. Intel.*, 2023, vol. 17, no. 4, pp. 2167–2191.
4. Orlov A.A., Abramova E.S. Development and research of a neural network alternate incremental learning algorithm. *Computer Optics*, 2023, vol. 47, no. 3, pp. 491–498.
5. Angelov P., Gu X. Autonomous learning multi-model classifier of 0-Order (ALMMo-0). 2017 *Evolving and Adaptive Intelligent Systems (EAIS)*, 2017, pp. 1–7.
6. Angelov P., Gu X., Principe J.C. Autonomous Learning Multimodel Systems from Data Streams. *IEEE Trans. Fuzzy Syst.*, 2018, vol. 26, no. 4, pp. 2213–2224.
7. Angelov P., Gu X., Kangin D. Empirical Data Analytics. *Int. J. Intell. Syst.*, 2017, vol. 32, no. 12, pp. 1261–1284.
8. Hochma Y., Last M. Fast online feature selection in streaming data. *Mach Learn.*, 2025, vol. 114, no. 1, pp. 1–35.
9. Angelov P. *Autonomous Learning Systems*. Hoboken, John Wiley and Sons, 2001. 224 p.
10. Moustafa N., Slay J. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). 2015 *Military Communications and Information Systems Conference (MilCIS)*, Canberra, Australia, 2015, pp. 1–6.

11. Vrbančič G., Fister I.Jr., Podgorelec V. Datasets for phishing websites detection. *Data in Brief.*, 2020, vol. 33, pp. 106438.

12. Spambase. Available at: <https://archive.ics.uci.edu/dataset/94/spambase> (Accessed: 02.10.2025).

13. James G., Witten D., Hastie R., Tibshirani R. Statistical Learning. *Springer Texts in Statistics*, 2021, pp. 15–57.

14. Din S.U., Shao J., Kumar J., Mawuli C.B. Data stream classification with novel class detection: a review, comparison and challenges. *Knowl. Inf. Syst.*, 2021, vol. 63, no. 9, pp. 2231–2276.

15. Losing V., Hammer B., Wersing H. Incremental on-line learning: A review and comparison of state of the art algorithms. *Neurocomputing*, 2018, vol. 275, pp. 1261–1274.

**Mikhail O. Svetlakov**

Candidate of Technical Sciences, Associate Professor of Department of Computer Systems in Control and Design, Tomsk State University of Control Systems and Radioelectronics (TUSUR)

40, Lenin pr., Tomsk, Russia, 634050

ORCID: 0000-0002-5060-0958

Phone: +7 (382-2) 41-47-17

Email: svetlakov.m4@gmail.com

**Igor G. Borovskoy**

Doctor of Physical and Mathematical Sciences, Head of the Department of Economic Mathematics, Informatics and Statistics, TUSUR

40, Lenin pr., Tomsk, Russia, 634050

ORCID: 0000-0003-2707-7203

Phone: +7 (382-2) 90-01-85

Email: igor.g.borovskoi@tusur.ru

Received: 11.10.2025.

Accepted: 13.01.2026.