

Интеграция вероятностной и нечеткой логики для улучшения интерпретации семантики естественного языка

Э.Р. Городецкий, Ю.М. Вишняков, Б.В. Белокобыльский

Кубанский государственный университет», Краснодар

Аннотация: Развитие цифровых платформ обучения, систем электронного документооборота, веб-ориентированных комплексов, работающих с текстовой информацией на естественном языке, повлекло за собой увеличению объема контента и/или массивов обрабатываемых полнотекстовых документов. Это, в свою очередь, актуализировало запрос на высокоэффективные методы обработки естественного языка, способные улавливать семантику текста. В статье предложена гибридная архитектура, основанная на интеграции вероятностной и нечеткой логики, которая эффективно решает проблемы семантической многозначности за счет интеграции стохастических и нечетко-логических каналов, учитывающих как статистические закономерности, так и лингвистическую неопределенность.

Ключевые слова: интеграция, семантика, интерпретация, естественный язык, неопределённость, закономерности.

Богатство и многозначность семантических структур естественных языков обуславливают необходимость интерпретационных подходов, способных учитывать градуированность смыслов и контекстную зависимость, что невозможно при использовании исключительно бинарной логики и формально-синтаксических моделей. Это стало стимулом к развитию семантической теории информации, которая проистекает из признания того факта, что эффективная коммуникация выходит за рамки надежной передачи символов и включает в себя понимание, интерпретацию и обмен значимыми знаниями. Признавая неоднозначность и непредсказуемость, которые неотъемлемо связаны с человеческим языком, существующие на сегодняшний день стратегии и подходы либо пытаются вычислить вероятность конкретных интерпретаций, либо функционируют в рамках степеней истинности [1].

В данном контексте следует отметить последние наработки в области вероятностных подходов и нечеткой логики. При использовании вероятностного подхода семантика рассматривается как азартная игра. Результаты всестороннего анализа данных используются для расчета

вероятности различных интерпретаций текста с учетом количества раз, когда эти интерпретации встречались в предыдущих ситуациях, аналогичных текущей. Эта стратегия может быть очень успешной при работе с огромными массивами данных, поскольку она позволяет обнаружить закономерности и связи, которые системы, основанные на правилах, могут упустить. С другой стороны, нечеткая логика черпает свои идеи из теории нечетких множеств и, в большей степени, ориентирована на приблизительное мышление, чем на точное соблюдение правил, как это характерно для традиционной логики [2]. В изучении семантики это означает, что на основе окружающей информации в предложении возможно множество интерпретаций высказывания.

В таблице 1 представлены результаты сравнительного анализа возможностей вероятностной и нечеткой логики для семантической интерпретации текстов.

Учитывая возможности и недостатки рассматриваемых методов, по мнению автора, значительный научный интерес приобретает объединение вероятностной и нечеткой логики для усиления интерпретации семантики естественного языка, что и предопределило выбор темы данной статьи.

Таблица 1. Сравнительная характеристика вероятностных и нечетких методов семантической интерпретации (составлено автором)

Аспект анализа	Вероятностные методы	Нечеткая логика	Следствие для семантического анализа
Источник неопределенности	Статистическая неопределенность, обусловленная ограниченностью и неоднородностью корпусов	Лингвистическая неопределенность, связанная с размытостью смысловых границ	Семантическая неоднозначность имеет смешанную природу и не может быть описана одним типом неопределенности
Чувствительность к полисемии	Высокая: вероятность распределяется между значениями, но без явной интерпретации переходов	Умеренная: допускаются частичные принадлежности, но без учета частотных закономерностей	Требуется совместный учет частот смыслов и степени их выраженности
Стабильность при смене контекста	Снижение точности на 10–20% при резком контекстном сдвиге	Колебания выходных оценок в пределах 15–25 % при изменении правил	Отсутствие механизма взаимной компенсации ошибок приводит к деградации качества
Интерпретируемость результатов	Ограниченная: вероятностные оценки трудно интерпретируемы лингвистически	Высокая: правила и функции принадлежности интерпретируемы экспертами	Интеграция позволяет связать численные оценки с лингвистическими категориями
Зависимость от объема обучающих данных	Существенная: снижение качества до 20 % при уменьшении корпуса	Незначительная: возможна работа на малых выборках	Комбинированный подход снижает зависимость от объема данных
Устойчивость к шуму и неточностям	Средняя: шум приводит к смещению распределений	Высокая: нечеткие правила сглаживают локальные искажения	Интеграция повышает общую робастность семантического вывода
Основное ограничение	Отсутствие формализации смысловой размытости	Отсутствие статистической адаптации	Ограничения методов являются взаимодополняющими

Над разработкой представления семантики естественного языка, которое использует вероятностную логику для интеграции логического и взвешенного неопределенного знания, трудятся В.В. Борисов, А.С. Федулов [3], Rashmi Thakur, Anil Vasoya, Manoj Chavan [4].

Возможности методов нечеткой логики, нашедших свое широкое применение в семантическом разборе, где они помогают отвечать на сложные запросы на естественном языке, требующие рассуждений и агрегации данных, описывают Sandro Pezzelle, Raquel Fernández [5], Чернышев А.Б. [6], Farimah

Houshmand-Nanehkaran, Seyed Mohammadreza [7].

Однако, хотя сегодня наблюдается активное развитие вероятностных и нечётких методов обработки естественного языка, их интеграция в единую семантическую модель остаётся недостаточно проработанной. К нерешённым проблемам относятся отсутствие общепринятого формального аппарата, обеспечивающего согласованное объединение вероятностной неопределённости и нечёткой семантической неоднозначностью, а также сложности интерпретации результатов гибридных моделей, особенно при работе с контекстно-зависимыми и прагматическими аспектами смысла [8].

Таким образом, цель статьи заключается в рассмотрении возможностей и особенностей интеграции вероятностной и нечеткой логики для улучшения интерпретации семантики естественного языка.

Итак, предполагается, что рассматриваемый в рамках данной статьи подход интеграции вероятностной и нечеткой логики, позволит объединить вычислительные возможности методов вероятностного анализа с инструментами нечеткой логики для повышения качества интерпретации семантики естественного языка. Основная цель такой интеграции — создавать более точные и информативные модели семантического анализа, одновременно учитывая языковую неоднозначность, контекстную зависимость и неопределённость — факторы, с которыми традиционные статистические и нейросетевые модели часто справляются недостаточно эффективно [9]. В таблице 2 представлены методические и математические предпосылки предлагаемой интеграции.

Таблица 2 Методические и математические предпосылки интеграции
вероятностных и нечётких методов (составлено автором)

Этап семантического анализа	Вероятностная формализация	Формализация на основе нечёткой логики	Предлагаемый интеграционный механизм
Представление семантических гипотез	Вероятностное распределение интерпретаций $P(s_i c), \sum_i P(s_i c) = 1$	Нечёткие множества семантических состояний с функциями принадлежности $\mu(s_i) \in [0,1]$	Совместное представление гипотезы в виде пары $\langle P(s_i c), \mu(s_i) \rangle$
Учёт контекста	Контекст моделируется через условные вероятности $P(s_i x, c)$ и скрытые переменные	Контекст задаётся через активацию правил $\mu_k(c) \cdot \mu(s_i)$	Контекст используется для одновременной корректировки $P(s_i)$ и $\mu(s_i)$
Обработка полисемии	Конкурирующие значения ранжируются по величине $P(s_i x)$	Допускается частичная принадлежность: $\exists_{s_i, s_j}: \mu(s_i) > 0, \mu(s_j) > 0$	Вероятностное ранжирование ограничивается нечётким порогом $\mu(s_i) \geq \alpha$
Агрегация признаков	Байесовская или линейная агрегация: $P(s_i x)$	Нечёткая агрегация: $\mu(s_i) = \max_k \min(\mu_k)$	Иерархическая схема: статистическая агрегация \rightarrow нечёткий вывод
Принятие семантического решения	Выбор максимальной апостериорной вероятности $s^* = \operatorname{argmax} P(s_i x)$	Выбор по максимальной степени принадлежности $s^* = \operatorname{argmax} \mu(s_i)$	Комбинированный критерий $s^* = \operatorname{argmax} P(s_i) \cdot \mu(s_i)$
Интерпретация результата	Числовая оценка $P(s_i)$ без лингвистического объяснения	Лингвистически интерпретируемые правила вида «ЕСЛИ–ТО»	Вероятностная оценка сопровождается объясняющим нечётким выводом
Обработка неопределённости	Стохастическая неопределённость, описываемая дисперсией $\operatorname{Var}(P)$	Лингвистическая размытость, моделируемая шириной функций принадлежности	Разделение стохастической и лингвистической неопределённости
Адаптация модели	Переобучение параметров $P^{t+1}(s_i) = P^t(s_i) + \Delta P$	Корректировка базы правил и функций $\mu^{t+1}(s_i)$	Вероятностная адаптация при фиксированной структуре нечётких правил
Вычислительная схема	Итеративные вычисления сложности ($O(n \times k)$)	Последовательный нечёткий вывод ($O(r \times k)$)	Двухуровневая схема без экспоненциального роста сложности

Принимая во внимание данные, приведенные в таблице 2, автором разработана гибридная архитектура, объединяющая вероятностный формализм и аппарат нечеткой логики, схематическое представление которой представлено на рис. 1. Такой подход позволяет не только количественно оценить правдоподобие гипотез, но и качественно моделировать семантические оттенки, приближая машинную интерпретацию к когнитивным процессам человека. Ключевым элементом предлагаемой системы является введение контура рекуррентной адаптации, обеспечивающего непрерывную калибровку параметров модели на основе минимизации функции потерь при выявлении онтологических коллизий.

Представленная на рис. 1 архитектура демонстрирует переход от линейной обработки текста к итеративному процессу смыслового синтеза. Декомпозиция вычислительного ядра на параллельные каналы — стохастический и нечетко-логический — позволяет системе эффективно работать с разнородными типами неопределенности [10].

Вероятностный канал минимизирует риск статистических ошибок при выборе значений полисемичных терминов, в то время как нечеткий контроллер обрабатывает качественные характеристики, недоступные для прямой векторизации.

Особую значимость представляет внедрение блока тензорной агрегации и онтологического валидатора. В отличие от стандартных методов, где выход нейросети принимается как данность, предложенная модель верифицирует полученный семантический граф на соответствие аксиоматике предметной области.

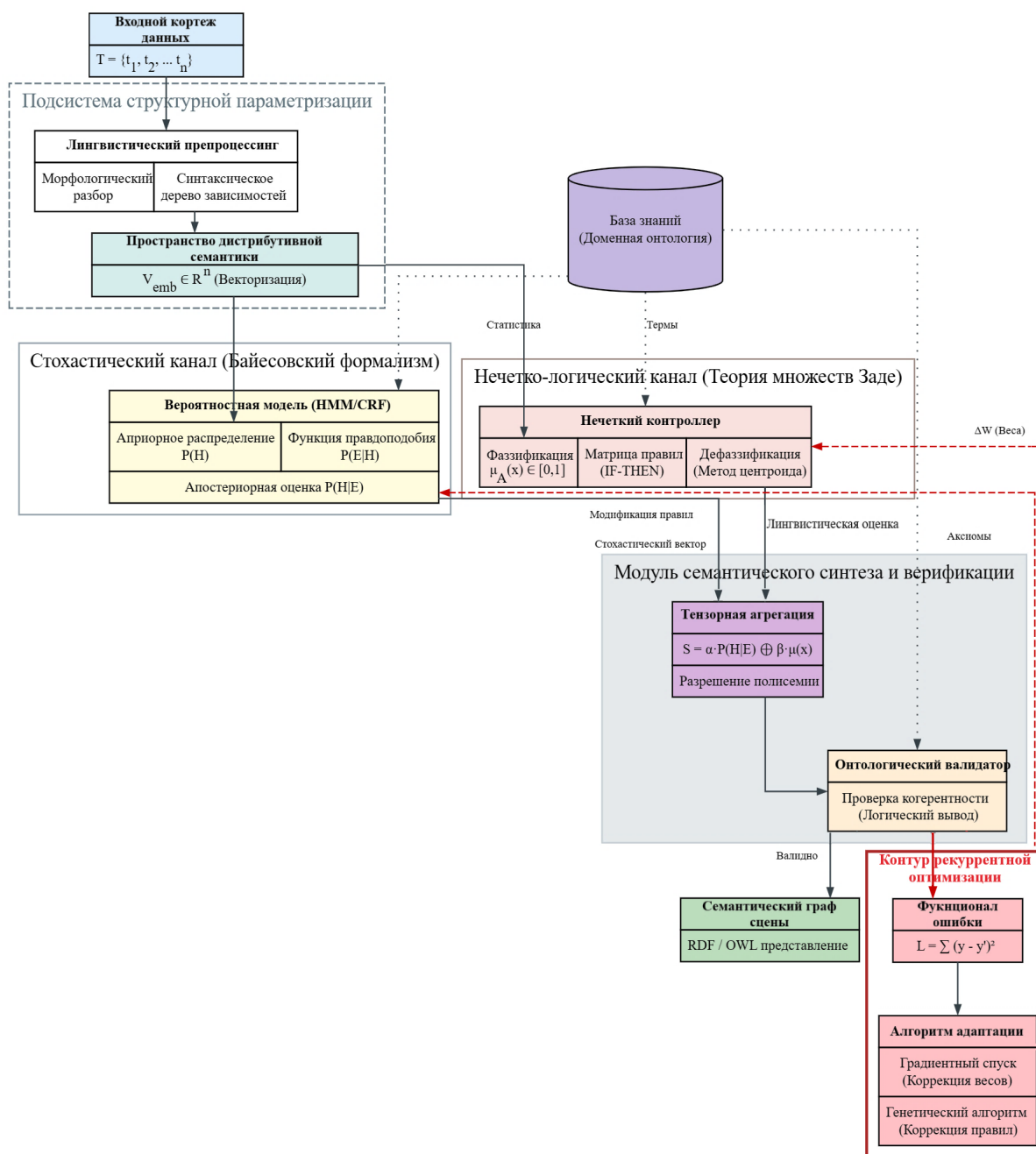


Рис. 1 Архитектура нейро-нечеткого вероятностного классификатора семантических отношений (составлено автором)

При обнаружении логических противоречий (коллизий) активируется контур рекуррентной оптимизации. Это позволяет динамически перестраивать весовые коэффициенты вероятностных моделей и адаптировать базу продукционных правил, тем самым реализуя принцип «обучения в процессе эксплуатации».

Математическая модель гибридного семантического вывода

Формализуем предлагаемую систему S как упорядоченный кортеж:

$$S = \langle T, V, \mathcal{M}_{prob}, \mathcal{M}_{fuzzy}, \Omega, \Psi, \mathcal{A} \rangle$$

где T — входное пространство текстовых данных; V — пространство векторных представлений (дистрибутивная семантика); \mathcal{M}_{prob} — стохастическая модель (вероятностный канал); \mathcal{M}_{fuzzy} — модель нечеткого логического вывода (нечеткий канал); Ω — доменная онтология (база знаний); Ψ — оператор гибридной агрегации; \mathcal{A} — механизм адаптации (контур обучения).

Пусть $t \in T$ — входная последовательность токенов $t = \{w_1, w_2, \dots, w_n\}$. Определим отображение ϕ (функцию векторизации), переводящую токены в d -мерное евклидово пространство:

$$\phi: T \rightarrow \mathbb{R}^{n \times d}, \quad x_i = \phi(w_i)$$

где $x_i \in V$ — вектор признаков i -го токена.

В рамках стохастического канала задача интерпретации сводится к поиску наиболее вероятной последовательности скрытых семантических состояний $H = \{h_1, \dots, h_n\}$ для наблюдаемой последовательности $X = \{x_1, \dots, x_n\}$. Используя байесовский формализм, апостериорную вероятность можно определить следующим образом:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

где, $P(H) = \prod_{i=1}^n P(h_i | h_{i-1})$ — априорная вероятность перехода (матрица переходов); $P(X|H) = \prod_{i=1}^n P(x_i | h_i)$ — функция правдоподобия (матрица эмиссии).

Выход стохастического канала — вектор вероятностей гипотез $p \in [0,1]^k$:

$$p = \text{softmax}(W_p \cdot h_{final} + b_p)$$

Определим лингвистическую переменную как $\langle \beta, U, \tilde{T}, G, M \rangle$. Входной

вектор x подвергается фаззификации через набор функций принадлежности $\mu_{A_j}(x)$:

$$\mu_{A_j}: U \rightarrow [0,1], \quad j = 1, \dots, m$$

База знаний содержит R продукционных правил следующего вида:

$$\text{Если } (x_1 \in \tilde{A}_1) \wedge \dots \wedge (x_d \in \tilde{A}_d) \text{ тогда } (y \in \tilde{B}_k)$$

Степень активации k -го правила вычисляется через T -норму (например, \min):

$$\alpha_k = \min_i (\mu_{A_{i,k}}(x_i))$$

Результирующее нечеткое множество выводится методом импликации (например, по Мамдани), а четкое значение y^* (дефаззификация) вычисляется методом центроида:

$$y^* = \frac{\int_Y y \cdot \mu_{agg}(y) dy}{\int_Y \mu_{agg}(y) dy}$$

Выход канала — вектор семантических оценок $f \in \mathbb{R}^k$.

Интеграция выходов двух каналов осуществляется оператором Ψ , реализующим взвешенную суперпозицию:

$$S = \Psi(p, f) = W_{agg} \cdot (\lambda \cdot p \oplus (1 - \lambda) \cdot \sigma(f))$$

где λ — динамический коэффициент доверия каналу, \oplus — оператор конкатенации или суммирования, σ — функция активации.

Пусть Ω задает множество аксиоматических ограничений C_Ω . Функция валидации v возвращает меру когерентности результата S :

$$v(S, \Omega) = \begin{cases} 1, & \text{если } \forall c \in C_\Omega: S \models c \\ 0, & \text{иначе (существует } c \in C_\Omega: S \not\models c) \end{cases}$$

где \models — отношение логического следования, гарантирующее непротиворечивость интерпретации S правилу c .

Задача обучения системы формулируется как минимизация функционала ошибки L при наличии онтологических ограничений. Целевая

функция потерь:

$$L(\theta) = L_{data}(S, y_{true}) + \gamma \cdot L_{ontology}(S, \Omega)$$

где $\theta = \{W_p, \mu, W_{agg}\}$ — множество обучаемых параметров.

Для стохастического канала обновление весов происходит методом градиентного спуска:

$$W_p^{(t+1)} = W_p^{(t)} - \eta \frac{\partial L}{\partial W_p}$$

Для нечеткого канала применяется коррекция параметров функций принадлежности (центров c и ширин σ гауссиан):

$$c_{new} = c_{old} - \eta \frac{\partial L}{\partial c}, \quad \sigma_{new} = \sigma_{old} - \eta \frac{\partial L}{\partial \sigma}$$

Таким образом, система S является динамической структурой, состояние, которое эволюционирует во времени t :

$$S(t+1) = \mathcal{A}(S(t), L(\theta))$$

Подводя итоги, отметим, что в статье описана разработанная автором гибридная архитектура нейро-нечеткого вероятностного классификатора семантических отношений, которая представляет собой перспективный подход к решению проблемы многозначности и контекстной зависимости в естественном языке. Предложенная архитектура наглядно демонстрирует механизм интеграции стохастического и нечетко-логического каналов обработки информации, что позволяет учитывать как статистические закономерности, так и лингвистическую неопределенность. Также формализована математическая модель авторского подхода, которая создает теоретический базис для реализации данной архитектуры в программных системах семантического анализа.

Ожидается, что интеграция вероятностных и нечетких методов в единый контур с адаптивным управлением позволит существенным образом повысить робастность системы интерпретации, обеспечивая высокую точность

извлечения знаний даже при работе с неструктурированными и зашумленными текстовыми массивами.

Литература

1. Ярушкина Н. Г., Мошкин В. С., Андреев И. А. Гибридный метод анализа тональности текстов на основе нечеткой логики и машинного обучения // Программные продукты и системы. 2019. № 4. С. 595–601.
2. Vashishtha S., Gupta V., Mittal M. Sentiment analysis using fuzzy logic: A comprehensive literature review // Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2023. Volume 13, Issue 5. pp. 20-29.
3. Борисов В.В. Федулов А.С. Методы и алгоритмы нечеткого поиска текстовой информации в семантических сетях // Вестник МЭИ. 2020. № 2. С. 112–119.
4. Thakur R., Vasoya A., Chavan M. Fuzzy-HAGRN: Fuzzy-Based Hierarchical Attention Gated Recurrent Network for Sentiment Classification and Review Rate Prediction // Computational Intelligence. 2025 .Volume 41, Issue 5. pp. 108-116.
5. Pezzelle S., Fernández R. Semantic Adaptation to the Interpretation of Gradable Adjectives via Active Linguistic Interaction // Cognitive Science. 2023. Volume 47, Issue 2. pp. 67-75.
6. Чернышев А. Б. Построение онтологии предметной области с использованием нечетких продукционных правил // Искусственный интеллект и принятие решений. 2022. № 3. С. 15–24.
7. Houshmand-Nanehkaran F., Mohammadreza Lajevardi S. Optimization of fuzzy similarity by genetic algorithm in user-based collaborative filtering recommender systems // Expert Systems. 2022. Volume 39, Issue 4. pp. 76-82.
8. Ульянов С.В., Мишин А.А. Квантовая нечеткая логика в задачах интеллектуального анализа семантической неопределенности // Системный

анализ: моделирование и управление. 2023. № 2. С. 89–98.

9. Глухих И.Н., Глухих К.И. Разработка экспертных систем на основе большой языковой модели и генерации с дополненной выборкой // Инженерный вестник Дона. 2025. № 11. URL: ivdon.ru/ru/magazine/archive/n11y2025/10496 (дата обращения: 18.12.2025).

10. Вакушин А.А., Клебанов Б.И. Применение больших языковых моделей в имитационном моделировании//Инженерный вестник Дона. 2024. №2. URL: ivdon.ru/ru/magazine/archive/n2y2024/8990 (дата обращения: 18.12.2025).

References

1. Yarushkina N. G., Moshkin V. S., Andreev I. A. Programmnye produkty i sistemy. 2019. No. 4. pp. 595–601.

2. Vashishtha S., Gupta V., Mittal M. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2023. Volume 13, Issue 5. pp. 20–29.

3. Borisov V. V., Fedolov A. S. Vestnik MEI. 2020. No. 2. pp. 112–119.

4. Thakur R., Vasoya A., Chavan M. Computational Intelligence. 2025. Volume 41, Issue 5. pp. 108-116.

5. Pezzelle S., Fernández R. Semantic Adaptation to the Interpretation of Gradable Adjectives via Active Linguistic Interaction // Cognitive Science. 2023. Volume 47, Issue 2. pp. 67-75.

6. Chernyshev A. B. Iskusstvennyj intellekt i prinyatie reshenij. 2022. No. 3. pp. 15-24.

7. Houshmand-Nanehkaran F., Mohammadreza Lajevardi S. Expert Systems. 2022. Volume 39, Issue 4. pp. 76–82.

8. Ulyanov S. V., Mishin A. A. Sistemnyj analiz: modelirovanie i upravlenie. 2023. No. 2. pp. 89–98.

9. Gluhih I.N., Gluhih K.I. Inzhenernyj vestnik Dona. 2025. № 11. URL: ivdon.ru/ru/magazine/archive/n11y2025/10496 (date assessed 18.12.2025).



10. Vakushin A.A., Klebanov B.I. Inzhenernyj vestnik Dona. 2024. №2.
URL: ivdon.ru/ru/magazine/archive/n2y2024/8990 (date assessed: 18.12.2025).

Дата поступления: 10.12.2025

Дата публикации: 7.02.2026