

Правило принятия решений группой искусственных интеллектуальных акторов «k-консенсус» на второй стадии метода ЭВОЛЮЦИОННОГО СОГЛАСОВАНИЯ

В.С. Качалин, В.И. Протасов

Аннотация — В статье представлено правило принятия решений группой искусственных интеллектуальных акторов «k-консенсус», предназначенное для второй стадии метода эволюционного согласования – этапа оценки ответов акторов. Разработанное правило позволяет выполнять анализ ответов нескольких классифицирующих моделей машинного обучения и на основе данного анализа формировать решение группы, которое, согласно теореме Кондорсе, имеет преимущество над индивидуальными решениями по точности и релевантности. Помимо положительного и отрицательного решения разработанное правило позволяет группе акторов сомневаться и отвечать «Не знаю», что подразумевает под собой передачу права вынесения решения человеку. Данный ответ обеспечивается за счет оценки решений акторов с помощью среднеквадратичного отклонения, энтропии Шеннона и генеральной средней метрик уверенности каждого актора. Данное правило имеет ограничение – оно может быть применено только для несоставных решений акторов. Выведены формулы для вычисления значений границ зоны неопределенности группы, а также представлено необходимое и достаточное условие существования данной области. Экспериментальная проверка разработанного правила показала, что оно значительно выигрывает у классических правил голосования в случаях, когда допускается отказаться от решения в спорных ситуациях. Проведено исследование чувствительности правила к изменению параметров, что показало его устойчивость и управляемость.

Ключевые слова — группа акторов, искусственный интеллект, метод эволюционного согласования, области решений, принятие решений, троичная логика, энтропия Шеннона.

I. ВВЕДЕНИЕ

Идея применения группы искусственных интеллектуальных акторов (под «искусственным интеллектуальным актором» в данной статье подразумевается обученная модель ИИ) не является принципиально новой: исследования по улучшению

результатов работы ИИ за счет объединения их в группы ведутся не первый год [1]–[3]. Увеличение точности и релевантности результатов группы акторов по сравнению с индивидуальными результатами отдельных моделей ИИ объясняется теоремой Кондорсе о жюри присяжных: вместе с увеличением количества акторов растет вероятность принятия правильного решения группой при условии, что каждый отдельный актер в группе принимает собственное правильное решение с вероятностью выше чем 0,5 [4]. Большим развитием концепции группового ИИ является Метод эволюционного согласования (МЭС) [5]. Характерной чертой МЭС является наличие стадий генерации ответа и экспертизы полученных ответов. На стадии генерации ответов искусственные интеллектуальные акторы на основе имеющейся информации принимают собственные решения. А на стадии экспертизы акторы обмениваются решениями и оценивают результаты друг друга, дополняя и улучшая свои собственные решения за счет правила взаимодействия, основанного на принципах генетических алгоритмов [6]. Однако данная особенность обладает некоторым недостатком – в существующем виде МЭС приспособлен к работе, если есть специальные акторы-эксперты, либо если производится переобучение акторов-генераторов, а это довольно долгий процесс. Акторы же, обученные как генераторы в существующем виде не могут работать, как эксперты. К тому же присутствует ограничение на применение МЭС: он хорошо подходит для слабо- или неструктурированных задач [7], а также когда решение обладает вариативностью, т.е. может видоизменяться полностью или частями и не принадлежит к ограниченному множеству решений с небольшим количеством возможных вариантов. Данная особенность делает избыточным и громоздким применение МЭС к задачам, требующим несоставное решение из сильно ограниченного множества вариантов: «Положительное решение», «Отрицательное решение» и «Неопределенность». Последний вариант обозначает воздержание от принятия какого-либо решения и является аналогом ответа «Не знаю». Целью данной работы как раз и является предложение альтернативного подхода к оценке решений акторов, и их преобразование в единый ответ всей группы искусственных интеллектуальных акторов для применения на второй

Статья получена 7 августа 2026 г.

В. С. Качалин, ФГАОУ ВО «Московский авиационный институт (национальный исследовательский университет)», Москва, РФ (e-mail: vasilii.kachalin@gmail.com)

В. И. Протасов, ФГАОУ ВО «Московский авиационный институт (национальный исследовательский университет)», Москва, РФ (e-mail: protonus@yandex.ru)

стадии метода эволюционного согласования, что позволяет расширить область применения МЭС на несоставные решения и упростить этап экспертизы решений с сохранением преимуществ данного метода.

II. ГРУППА ИСКУССТВЕННЫХ ИНТЕЛЛЕКТУАЛЬНЫХ АКТОРОВ

Группа искусственных интеллектуальных акторов состоит из нескольких обученных классифицирующих моделей машинного обучения, которые получают на вход факторы и ограничения, на основе которых необходимо принять решение и на выходе дают свои собственные решения. На основе данных собственных решений необходимо принять общее – групповое решение.

Выбор моделей, относящихся к классифицирующим, обуславливается тем что задача принятия решения положительного или отрицательного решения, может рассматриваться как задача классификации, т.к. по своей сути она отображает множество факторов и ограничений во множество решений (положительные и отрицательные), а, следовательно, для решения этой задачи в качестве искусственных интеллектуальных акторов следует применять методы машинного обучения, которые ориентированы на классификацию объектов. При этом модели ИИ не обязательно должны быть одного вида, т.е. в одной группе могут быть дерево решений, наивный байесовский классификатор и многослойный персептрон.

Стоит также отметить, что акторы не должны являться копиями друг друга, потому что в таком случае они будут одинаковым образом анализировать факторы и ограничения и придут к одному и тому же результату одним и тем же путем, из-за чего эффективность данной группы будет равносильна применению одного такого актора и никакого качественного улучшения результатов наблюдаться не будет.

III. ФОРМИРОВАНИЕ ПРАВИЛА «К-КОНСЕНСУС»

При формировании правила, по которому группа искусственных интеллектуальных акторов будет принимать решения, стоит отталкиваться от принятия решений большинством голосов, т.е. когда положительное решение принимается в случае, когда за него проголосовало минимум $\frac{N}{2} + 1$, где N – общее количество акторов в группе. Графически (на рис. 1) данная ситуация представляет собой отрезок длиной N на числовой прямой, разделенный пополам. Данное деление образует две области на отрезке: $[0; \frac{N}{2}]$ и $[\frac{N}{2} + 1; N]$, первая область обозначает групповое отрицательное решение, а вторая – групповое положительное решение. Предположим, что m – количество акторов проголосовавших положительно и его значение, т.е. попадание в ту или иную область обозначает соответствующее групповое решение, в таком случае правило большинства можно представить в виде формулы:

$$D(x, y) = \begin{cases} OP, & 0 \leq m \leq \frac{N}{2} \\ PP, & \frac{N}{2} + 1 \leq m \leq N \end{cases},$$

где $D(x, y)$ – решение группы акторов; x – множество факторов; y – множество ограничений; OP – отрицательное решение; PP – положительное решение; m – количество акторов, принявших положительное решение; N – общее количество акторов в группе.

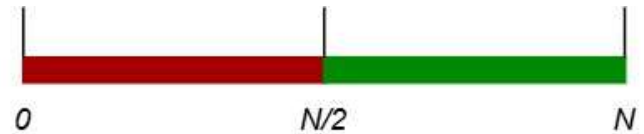


Рис. 1. Отрезок, графически отображающий области отрицательных (красный) и положительных (зеленый) решений группы, состоящей из N акторов

Очевидно, что для введения третьего ответа «Не знаю», отображающего неопределенность группы, необходимо добавить еще одну область, которая будет расположена между областями отрицательных и положительных решений. Чтобы определить границы новой области необходимо ввести две константы $k_{нижн}$ и $k_{верх}$, которые будут отображать значения, отделяющие области друг от друга. Таким образом образуются три области: $[0; k_{нижн}]$, $(k_{нижн}; k_{верх})$, $[k_{верх}; N]$. Первая область – область отрицательных решений, вторая область – область неопределенности, третья – область положительных решений. Получившийся отрезок с тремя областями представлен на рис. 2.



Рис. 2. Отрезок с тремя областями: отрицательных (красный), неопределенных (голубой) и положительных (зеленый) решений группы, состоящей из N акторов

В виде формулы данный подход выглядит следующим образом:

$$D(x, y) = \begin{cases} OP, & 0 \leq m \leq k_{нижн} \\ HЗ, & k_{нижн} < m < k_{верх} \\ PP, & k_{верх} \leq m \leq N \end{cases},$$

где $D(x, y)$ – решение группы акторов; x – множество факторов; y – множество ограничений; OP – отрицательное решение; $HЗ$ – решение, обозначающее неопределенность; PP – положительное решение; m – количество акторов, принявших положительное решение; $k_{нижн}$ – нижний порог, отделяющий область

отрицательных решений от области неопределенности; $k_{\text{верх}}$ – верхний порог, отделяющий область неопределенности от области положительных решений; N – общее количество акторов в группе.

Стоит отметить частный случай, когда $k_{\text{нижн}} = 0$ и $k_{\text{верх}} = N$. В такой ситуации данное правило будет формировать конкретный ответ только в случае единогласного решения всеми акторами, в противном случае за групповое решение принимается ответ «Не знаю». Такая вариация граничных значений позволяет группе акторов избегать любого риска и может использоваться, когда цена ошибки слишком высока.

Таким образом, верно и другое утверждение, что, варьируя значения $k_{\text{нижн}}$ и $k_{\text{верх}}$, можно регулировать склонность группы акторов к риску – чем больше область неопределенности, тем меньше вероятность группы принять неверное решение, и наоборот – чем меньше данная область, тем группа более склонна к риску. Данный момент стоит учитывать при формировании области неопределенности, чтобы найти оптимальный размер данной области для конкретной задачи, когда цена ответа «Не знаю» значительно ниже стоимости ошибки, и при этом доля таких ответов не превышает допустимой, чтобы избежать случая, когда группа всегда отвечает «Не знаю».

Для большей объективности следует ввести в правило оценку решения, как это делается в МЭС. Однако, в отличие от метода эволюционного согласования, оценивать решения стоит другим образом нежели применяя самих акторов для этой цели. Это позволит избежать потенциальных временных и вычислительных затрат.

Для этого вместо количества положительных решений акторов следует учитывать их метрику уверенности в положительном решении, а точнее генеральную среднюю метрику уверенности. Это позволяет оценивать среднее квадратичное отклонение, которое показывает степень разброса уверенностей акторов по принятию положительного решения. Соответственно, если СКО превышает определенный допустимый порог, то независимо от значения генеральной средней уверенности решением группы считается ответ, символизирующий неопределенность. Таким образом правило принятия решений принимает следующий вид:

$$D(x, y) = \begin{cases} \text{ОР}, & 0 \leq \bar{\mu} \leq k_{\text{нижн}}, \quad \sigma < \sigma_{\text{max}} \\ \text{НЗ}, & k_{\text{нижн}} < \bar{\mu} < k_{\text{верх}} \vee \sigma \geq \sigma_{\text{max}}, \\ \text{ПР}, & k_{\text{верх}} \leq \bar{\mu} \leq 1, \quad \sigma < \sigma_{\text{max}} \end{cases}$$

$$\bar{\mu} = \frac{\sum_{i=1}^N \mu_i}{N}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^N \mu_i^2}{N} - \left(\frac{\sum_{i=1}^N \mu_i}{N} \right)^2}$$

где $D(x, y)$ – решение группы акторов; x – множество факторов; y – множество ограничений; ОР – отрицательное решение; НЗ – решение, обозначающее неопределенность; ПР – положительное решение; $\bar{\mu}$ – генеральная средняя метрику уверенности акторов в положительном решении; μ_i – значение метрики уверенности i -го актора в положительном решении; $k_{\text{нижн}}$ – нижний порог, отделяющий область отрицательных решений от области неопределенности; $k_{\text{верх}}$ – верхний порог, отделяющий область неопределенности от области положительных решений; σ – среднее квадратичное отклонение уверенностей акторов в положительном решении; σ_{max} – максимально допустимое значение для среднее квадратичного отклонения уверенностей акторов в положительном решении; N – общее количество акторов в группе.

Т.к. СКО измеряет величину разброса уверенностей акторов относительно среднего, то может возникнуть ситуация, при которой уверенности акторов равномерно распределены по всему отрезку $[0, 1]$, в таком случае принятие конкретного решения чревато издержками, т.к. перевеса в какую-то сторону может и не быть. При этом аналогичное значение СКО может получиться и при подавляющим количестве уверенностей акторов в конкретной зоне при сохранении позиций крайних уверенностей акторов (хотя бы по одной с каждой стороны), что в свою очередь уже должно говорить о принятии более конкретного решения. Исходя из этого имеет смысл ввести дополнительную метрику, на которую можно полагаться при принятии решения группой акторов. В качестве такой метрики предлагается использовать энтропию Шеннона, которая в данном случае будет реагировать на распределение уверенностей акторов. Сама энтропия Шеннона для данной задачи вычисляется по следующей формуле:

$$H = -\bar{\mu} \log_2 \bar{\mu} - (1 - \bar{\mu}) \log_2 (1 - \bar{\mu}),$$

где H – энтропия Шеннона, отображающая несогласованность акторов, $\bar{\mu}$ – генеральная средняя метрику уверенности акторов в положительном решении.

Значение метрики варьируется в промежутке $[0, 1]$, при том что 1 обозначает «хаос», т.е. полную несогласованность акторов, а 0 – отсутствие «хаоса» – полную согласованность акторов.

Очевидно, что само значение энтропии должно сравниваться с некоторым порогом H_{max} , который бы характеризовал допустимую несогласованность. Таким образом если значение метрики выше этого порога, то ответом группы считается «Не знаю», в противном случае выполняется проверка СКО, которая в свою очередь определяет расхождение уверенностей относительно максимально допустимого значения; и затем проверяется генеральная средняя уверенностей относительно нижнего и верхнего порогов. Таким образом правило принятия решений группой искусственных интеллектуальных акторов «к-консенсус» принимает окончательный вид:

$$D = \begin{cases} OP, & 0 \leq \bar{\mu} \leq k_{\text{нижн}}, \sigma < \sigma_{\text{max}}, H < H_{\text{max}} \\ HЗ, & k_{\text{нижн}} < \bar{\mu} < k_{\text{верх}} \vee \sigma \geq \sigma_{\text{max}} \vee H \geq H_{\text{max}} \\ PR, & k_{\text{верх}} \leq \bar{\mu} \leq 1, \sigma < \sigma_{\text{max}}, H < H_{\text{max}} \end{cases}$$

$$\bar{\mu} = \frac{\sum_{i=1}^N \mu_i}{N}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^N \mu_i^2}{N} - \left(\frac{\sum_{i=1}^N \mu_i}{N} \right)^2}$$

$$H = -\bar{\mu} \log_2 \bar{\mu} - (1 - \bar{\mu}) \log_2 (1 - \bar{\mu})$$

где D – решение группы акторов; ОП – отрицательное решение; НЗ – решение, обозначающее неопределенность; ПР – положительное решение; $\bar{\mu}$ – генеральная средняя метрик уверенности акторов в положительном решении; μ_i – значение метрики уверенности i -го актора в положительном решении; $k_{\text{нижн}}$ – нижний порог, отделяющий область отрицательных решений от области неопределенности; $k_{\text{верх}}$ – верхний порог, отделяющий область неопределенности от области положительных решений; σ – среднее квадратичное отклонение уверенностей акторов в положительном решении; σ_{max} – максимально допустимое значение для среднее квадратичного отклонения уверенностей акторов в положительном решении; H – энтропия Шеннона, отображающая несогласованность акторов; H_{max} – пороговое значение для энтропии Шеннона; N – общее количество акторов в группе.

Таким образом, разработано параметризованное пороговое правило принятия решений группой искусственных интеллектуальных акторов «к-консенсус».

IV. ОПРЕДЕЛЕНИЕ ЗНАЧЕНИЙ ГРАНИЦ В ПРАВИЛЕ «К-КОНСЕНСУС» НА ОСНОВЕ БАЙЕСОВСКОГО ПОДХОДА

Т.к. размеры области неопределенности влияют на эффективность принятия решений группой акторов, то значения границ этой области следует выбирать исходя из возможных потерь, которые могут повлечь за собой неправильное решение или ответ «Не знаю».

Возможные релевантные решения могут быть двух видов: целесообразно положительное решение и целесообразно отрицательное решение:

$$\theta \in \{0, 1\},$$

где θ – релевантное решение; 0 – целесообразно отрицательное решение; 1 – целесообразно положительное решение.

Группа акторов может выдать один из нескольких ответов: положительное решение, отрицательное решение, «Не знаю». Исходя из этого можно сформировать матрицу потерь (L) при различных сочетаниях релевантного (θ) и принятого решения (D). Данная матрица представлена в табличной форме в

таблице I.

Таблица I. Матрица потерь при различных сочетаниях релевантного и принятого решения

$\theta \setminus D$	Отрицательное решение	«Не знаю»	Положительное решение
0	0	$C_{\text{НЗ}}$	$C_{\text{ЛП}}$
1	$C_{\text{ЛО}}$	$C_{\text{НЗ}}$	0

В данной матрице $C_{\text{ЛП}}$ – стоимость ложноположительного решения, $C_{\text{ЛО}}$ – стоимость ложноотрицательного решения, $C_{\text{НЗ}}$ – стоимость неопределенности (ответ «Не знаю»); все стоимости имеют значение больше нуля. Стоит сделать предположение, что неопределенность обходится дешевле, чем любая из ошибок, иначе смысл ответа «Не знаю» теряется:

$$0 < C_{\text{НЗ}} < \min(C_{\text{ЛП}}, C_{\text{ЛО}}),$$

где $C_{\text{НЗ}}$ – стоимость неопределенности; $C_{\text{ЛП}}$ – стоимость ложноположительного решения; $C_{\text{ЛО}}$ – стоимость ложноотрицательного решения.

Предположим, что известны: априорная вероятность релевантности положительного решения, априорная вероятность релевантности отрицательного решения и условные плотности распределения средней уверенности при каждом релевантном решении. Тогда апостериорная вероятность того, что релевантным решением будет положительное, вычисляется по формуле Байеса:

$$p = P(\theta = 1 | \bar{\mu}) = \frac{f(\bar{\mu} | \theta = 1) P(\theta = 1)}{f(\bar{\mu} | \theta = 1) P(\theta = 1) + f(\bar{\mu} | \theta = 0) P(\theta = 0)},$$

где $P(\theta = 1 | \bar{\mu})$ – апостериорная вероятность релевантности положительного решения; $f(\bar{\mu} | \theta = 1)$ – условная плотность распределения средней уверенности при релевантном положительном решении; $P(\theta = 1)$ – априорная вероятность релевантности положительного решения; $f(\bar{\mu} | \theta = 0)$ – условная плотность распределения средней уверенности при релевантном отрицательном решении; $P(\theta = 0)$ – априорная вероятность релевантности отрицательного решения.

Теперь можно вывести формулу для условного байесовского риска. Так для каждого возможного решения $D(x, y)$ условный риск при наблюдаемом $\bar{\mu}$ определяется как математическое ожидание потерь по апостериорному распределению:

$$R(D | \bar{\mu}) = \sum L(\theta, D) P(\theta | \bar{\mu}),$$

где $R(D|\bar{\mu})$ – условный риск для каждого возможного решения; $L(\theta, D)$ – матрица потерь; $P(\theta|\bar{\mu})$ – апостериорная вероятность релевантности положительного или отрицательного решения;

Раскрывая выражение для возможных решений, получаем:

$$R(OP|\bar{\mu}) = L(0, OP) * (1-p) + L(1, OP) * p = C_{ЛО} * p$$

$$R(PP|\bar{\mu}) = L(0, PP) * (1-p) + L(1, PP) * p = C_{ЛП} * (1-p)$$

$$R(HЗ|\bar{\mu}) = L(0, HЗ) * (1-p) + L(1, HЗ) * p = C_{HЗ}$$

Оптимальное решение выбирается по критерию минимума условного риска на основе оптимального байесовского решающего правила:

$$D^* = \arg \min_{D \in \{OP, HЗ, PP\}} R(D|\bar{\mu}),$$

где D^* – оптимальное решение; $R(D|\bar{\mu})$ – условный риск для каждого возможного решения.

Необходимо попарно сравнить условные риски. Отрицательное решение предпочтительнее положительного в случае, если:

$$C_{ЛО} * p < C_{ЛП} * (1-p) \\ \Downarrow \\ p < \frac{C_{ЛП}}{C_{ЛП} + C_{ЛО}}$$

где p – апостериорная вероятность релевантности положительного решения; $C_{ЛП}$ – стоимость ложноположительного решения; $C_{ЛО}$ – стоимость ложноотрицательного решения.

Обозначим выражение из правой части неравенства как k_{cp} :

$$k_{cp} = \frac{C_{ЛП}}{C_{ЛП} + C_{ЛО}},$$

Полученное выражение можно рассматривать как порог для бинарного случая, когда отсутствует область неопределенности. Данный порог k_{cp} делит область решений на области положительных и отрицательных решений.

Отрицательное решение предпочтительнее неопределенности в случае, если:

$$C_{ЛО} * p < C_{HЗ} \\ \Downarrow \\ p < \frac{C_{HЗ}}{C_{ЛО}}$$

где p – апостериорная вероятность релевантности положительного решения; $C_{HЗ}$ – стоимость неопределенности; $C_{ЛО}$ – стоимость ложноотрицательного решения.

Аналогично, правая часть неравенства представляет собой один из порогов, в данном случае, т.к. сравнивались отрицательное решение и неопределенность – нижнюю границу:

$$k_{нижн} = \frac{C_{HЗ}}{C_{ЛО}},$$

Положительное решение предпочтительнее неопределенности в случае, если:

$$C_{ЛП} * (1-p) < C_{HЗ} \\ \Downarrow \\ p > 1 - \frac{C_{HЗ}}{C_{ЛП}}$$

где p – апостериорная вероятность релевантности положительного решения; $C_{HЗ}$ – стоимость неопределенности, $C_{ЛП}$ – стоимость ложноположительного решения.

Как и в двух предыдущих случаях, правая часть неравенства представляет собой порог, в данном случае – верхнюю границу:

$$k_{верх} = 1 - \frac{C_{HЗ}}{C_{ЛП}},$$

Для того, чтобы существовало три области: отрицательные решения, неопределенность и положительные решения, необходимо выполнение неравенства:

$$k_{нижн} < k_{верх},$$

Раскрывая данные пороги получаем:

$$\frac{C_{HЗ}}{C_{ЛО}} < 1 - \frac{C_{HЗ}}{C_{ЛП}} \\ \Downarrow \\ C_{HЗ} < \frac{C_{ЛП} C_{ЛО}}{C_{ЛП} + C_{ЛО}}$$

где $C_{нз}$ – стоимость неопределенности; $C_{лп}$ – стоимость ложноположительного решения; $C_{ло}$ – стоимость ложноотрицательного решения.

Данное условие является необходимым и достаточным для существования области неопределенности. Если же стоимость неопределенности становится больше значения выражения из правой части, то область неопределенности исчезает и правило «k-консенсус» следует применять в бинарном варианте, когда есть только положительное и отрицательное решение. В таком случае в качестве границы между областями выступает значение $k_{ср}$.

Т.к. данные формулы для порогов выведены через апостериорную вероятность необходимости ремонта и стоимости ошибок, то чтобы с ними сравнивать генеральную среднюю уверенностей моделей, необходимо сделать допущение, что акторы являются идеально откалиброванными, откуда следует возможность использования вместо апостериорной вероятности генеральную среднюю уверенностей акторов.

V. ЭКСПЕРИМЕНТАЛЬНАЯ ПРОВЕРКА ПРАВИЛА «K-КОНСЕНСУС»

С целью проверки разработанного правила был проведен соответствующий эксперимент. Для этого был взят публичный датасет «Rain in Australia» [8]. Данный датасет содержит в себе ежедневные данные метеорологических наблюдений за 10 лет. Задача при работе с данным датасетом – предсказать будет ли «завтра» дождь на основе «сегодняшних» показателей. Датасет состоит из 142 193 строчек, пригодных к применению, т.е. в которых присутствует целевой показатель. На этапе предобработки числовые пропуски заполнялись медианой, а категориальные пропуски – модой, также был удален столбец с датой. Сам датасет был разбит на обучающую и тестовую выборки размерами 113 754 и 28 439 строчек соответственно.

Группа акторов была сформирована из четырнадцати случайных моделей машинного обучения: метод опорных векторов, многослойный перцептрон, дерево решений, логистическая регрессия, наивный байесовский классификатор, линейный и квадратичный дискриминантный анализ. Каждой модели в группе было по два экземпляра. Каждая модель обучалась на своей подвыборке, чтобы снизить корреляцию между моделями.

В качестве границ $k_{нижн}$ и $k_{верх}$ были приняты значения 0,4 и 0,6 соответственно; $\sigma_{тах} = 0,25$; $H_{тах} = 0,9$.

Непосредственно для проведения эксперимента была написана программа на языке программирования Python версии 3.10.11 с использованием библиотеки машинного обучения scikit-learn версии 1.5.2.

В качестве метрик в эксперименте применялись Accuracy Rate и его модификация.

Стандартный Accuracy Rate показывает отношение количества релевантных решений к общему числу всех рассмотренных наблюдений. Для более точной картины была произведена его модификация, а именно

отношение релевантных решений к общему числу всех рассмотренных наблюдений за исключением ситуаций, в которых группа выдала ответ «Не знаю». Данная модификация метрики позволяет оценить «чистую» точность группы акторов без учета ответа, подразумевающего неопределенность, т.е. как часто группа не ошибается.

Для сравнительного анализа с существующими методами голосования данная группа дополнительно принимала решения на основе подходов Hard Voting (голосование большинством голосов) и Soft Voting (голосование на основании усреднения вероятностей). Отдельно стоит сказать, что для данных подходов Accuracy Rate и его модифицированная версия имеют одно и то же значение, т.к. они не дают ответа «Не знаю».

Результаты эксперимента представлены в таблице II.

Таблица II. Результаты эксперимента проверки разработанного правила

	k-консенсус	Hard Voting	Soft Voting
Релевантные решения	20 833	22 707	22 648
Ответы «Не знаю»	5 498	0	0
Нерелевантные решения	2 108	5 732	5 791
Accuracy Rate	0,733	0,798	0,796
Модификация Accuracy Rate	0,908	0,798	0,796

Из представленных результатов видно, что количество релевантных решений у правила «k-консенсус» меньше, чем у классических подходов к голосованию на 6,6% от общего размера тестовой выборки. Это объясняется наличием области неопределенности, в которую попала часть релевантных ответов. Однако, в эту область также попала часть нерелевантных ответов, что позволило воздержаться от ответа при отсутствии уверенности в его релевантности.

Метрика Accuracy Rate у разработанного правила имеет значение 0,733 против 0,798 и 0,796 у Hard Voting и Soft Voting соответственно, что свидетельствует о некотором проигрыше данного правила в случае, если необходимо однозначно принять решение и отказ от ответа воспринимается как нерелевантное решение.

Однако, модифицированная метрика Accuracy Rate имеет значение 0,908, что демонстрирует преимущество правила перед классическими подходами из-за возможности отказаться от заведомо нерелевантного ответа. Данный показатель имеет большую важность в случаях, когда цена ошибки слишком высока.

Полученные показатели демонстрируют, что когда предотвращение ошибок играет важную роль и можно допустить ответ группы как неопределенность, то данное правило показывает очень хорошие результаты.

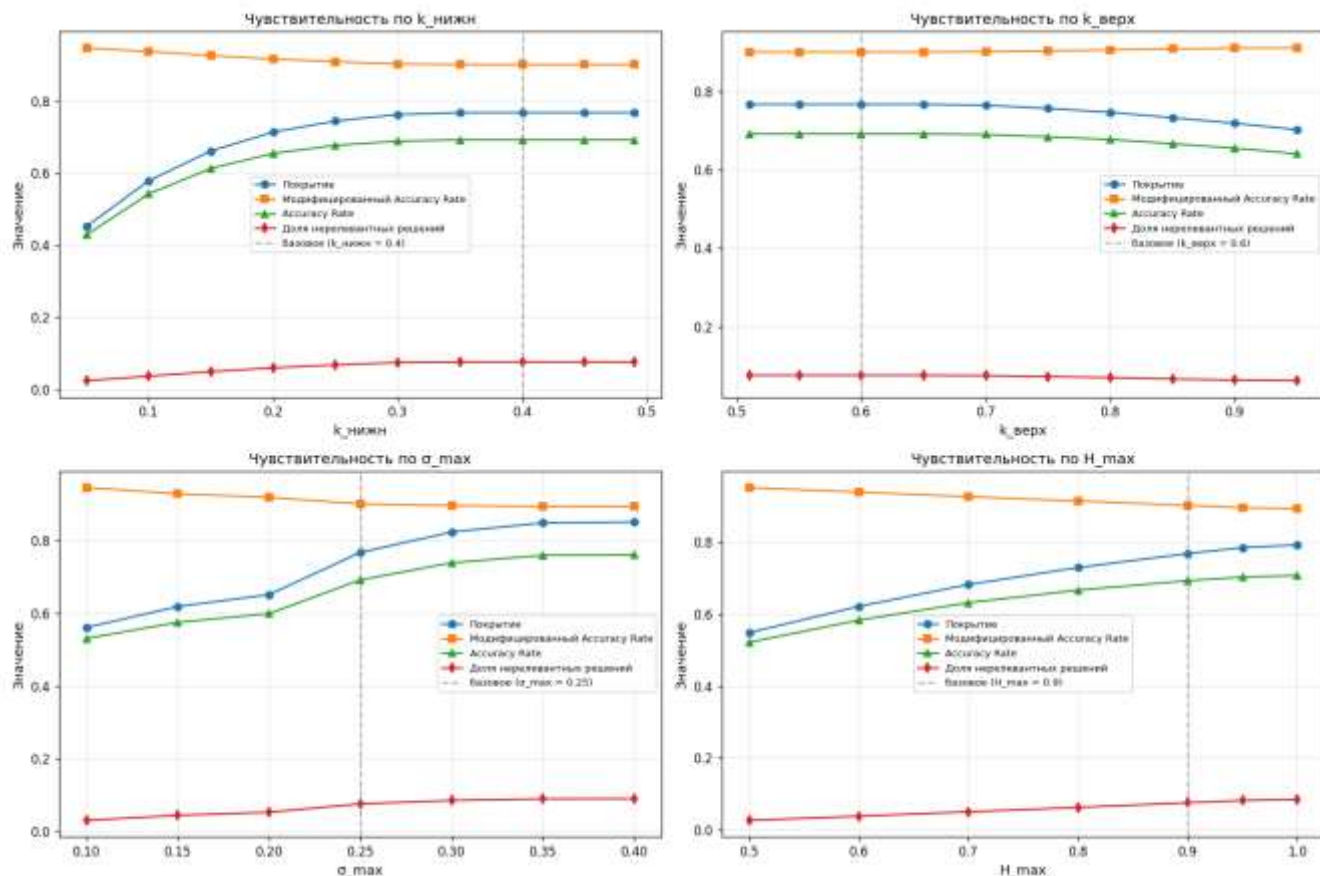
VI. ИССЛЕДОВАНИЕ ЧУВСТВИТЕЛЬНОСТИ ПРАВИЛА «K-КОНСЕНСУС»

Отдельно было проведено исследование чувствительности разработанного правила относительно изменяемых параметров: $k_{нижн}$, $k_{верх}$, σ_{max} и H_{max} . Для этого на основе датасета из предыдущего раздела был проведен соответствующий эксперимент. В рамках исследования фиксировались следующие метрики: покрытие (кол-во конкретных ответов), Accurasy Rate, модифицированный Accurasy Rate, доля нерелевантных решений.

Исследование было проведено в два этапа. На первом

остальные были зафиксированы на базовых значениях. На втором этапе исследовалось изменение одновременно двух параметров при фиксации оставшихся. Базовые значения параметров: $k_{нижн} = 0,4$; $k_{верх} = 0,6$; $\sigma_{max} = 0,25$; $H_{max} = 0,9$.

По результатам первого этапа были построены графики чувствительности разработанного правила к изменению параметров по одному (рис. 3) и график, отображающий компромисс между показателями покрытие и модифицированный Accurasy Rate (рис. 4).



этапе изменялся только один параметр, в то время как

Рис. 3. Графики чувствительности правила «к-консенсус» к изменению параметров. Покрытие (синий), Accurasy Rate (зеленый), модифицированный Accurasy Rate (оранжевый), доля нерелевантных решений (красный)

На графиках чувствительности по $k_{нижн}$ и $k_{верх}$ наблюдается асимметрия, это объясняется классовой асимметрией датасета. График чувствительности по $k_{нижн}$ при увеличении значения параметра показывает небольшое ухудшение по метрикам: модифицированный Accurasy Rate немного проседает, при одновременном незначительном росте доли нерелевантных решений. В то же время растет покрытие и стандартный Accurasy Rate. Это объясняется уменьшением зоны неопределенности – группа чаще дает определенный ответ нежели воздерживается от него. При этом при достижении параметром значения 0,35 метрики переходят в плато – дальнейшее увеличение параметра не приводит к каким-либо изменениям.

График чувствительности по $k_{верх}$ показывает обратное, что логично, т.к. увеличение его значения приводит к увеличению зоны неопределенности. Однако скос метрик в правой части незначителен по сравнению с поведением наблюдаемом при изменении нижней границы области неопределенности.

График чувствительности по σ_{max} демонстрирует не такую плавную картину. При ослаблении ограничения допустимого разброса ответов группы происходит резкий скачок на промежутке 0,2 – 0,3. Ухудшается модифицированный Accurasy Rate, а все остальные метрики растут, включая долю нерелевантных решений.

Более плавное изменение метрик демонстрируется на графике чувствительности по H_{max} . Здесь аналогично – ослабление ограничения ведет к ухудшению модифицированного Accurasy Rate при увеличении значений остальных метрик.

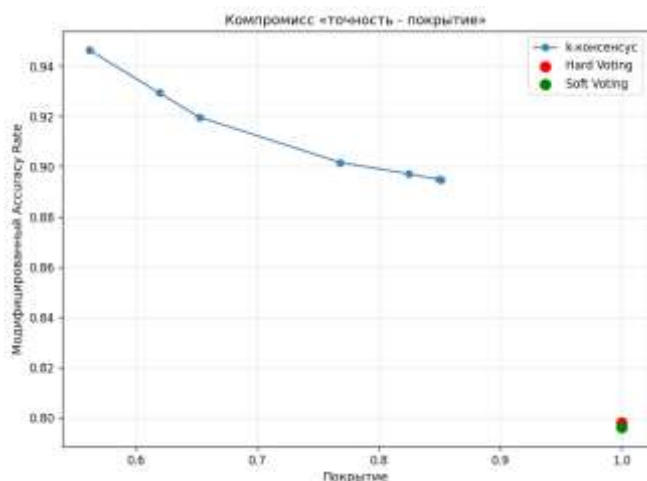


Рис. 4. Компромисс между покрытием и модифицированным Accuracy Rate с Hard Voting (красный) и Soft Voting (зеленый)

График, отображающий компромисс между покрытием и модифицированным Accuracy Rate, демонстрирует особенность правила – увеличение одного показателя влечет за собой уменьшение другого. Стоит отметить, что график обрывается на покрытии в 0,85, данный факт обусловлен тем что при построении графика в этой точке правило упиралось в ограничение по $k_{нижн}$ и $k_{верх}$. Если же убрать это и другие ограничения, то правило преобразуется в Soft Voting и будет иметь ее характеристики. Также наличие Hard Voting и Soft Voting на графике демонстрирует преимущество разработанного правила при готовности к допущению снижения покрытия.

По результатам второго этапа были построены тепловые карты чувствительности по каждой метрике при одновременном изменении σ_{max} и H_{max} (рис. 5).

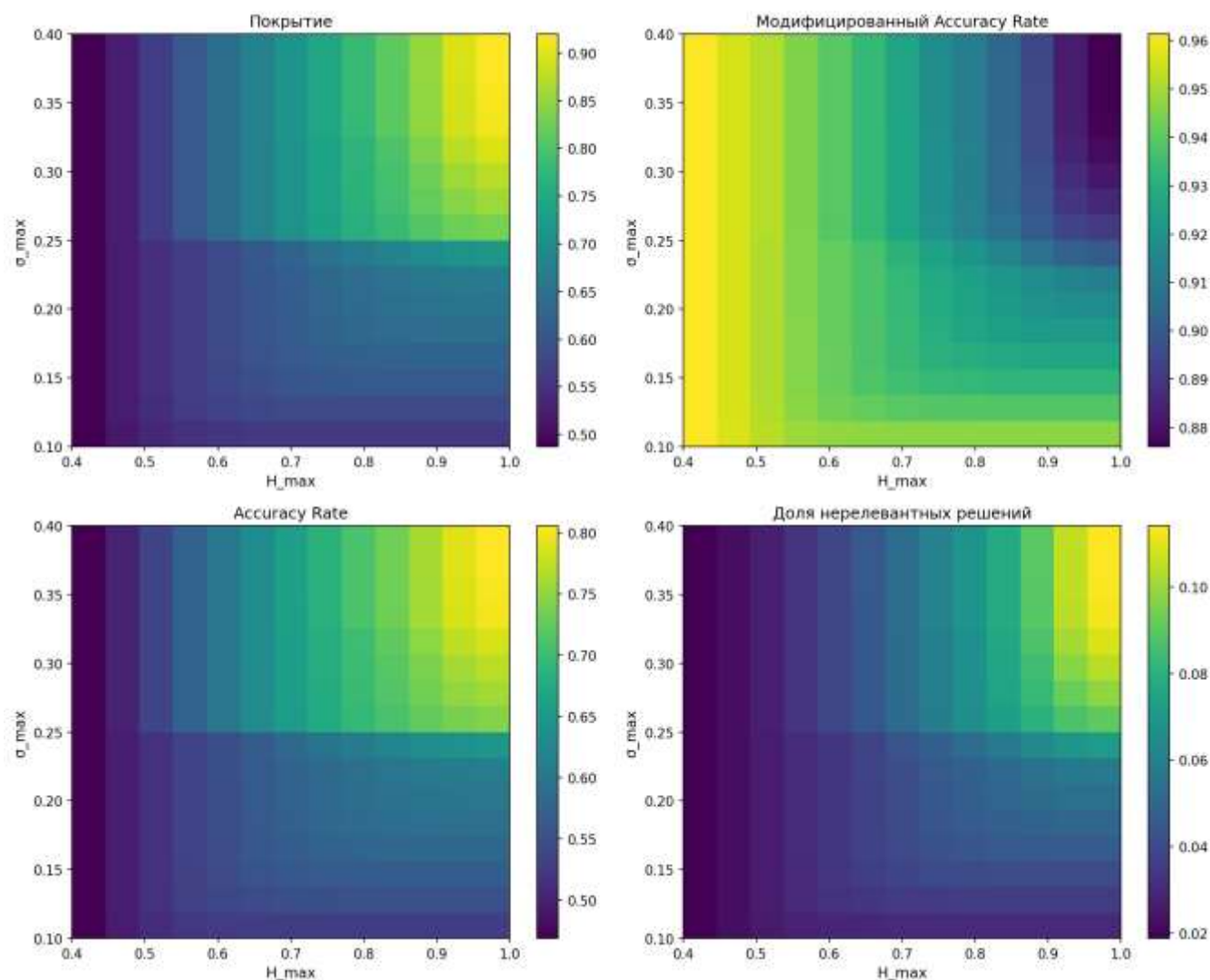


Рис. 5. Тепловые карты чувствительности правила «k-консенсус» по каждой метрике при одновременном изменении σ_{max} и H_{max}

Данные тепловые карты демонстрируют резкий скачок по всем метрикам при $\sigma_{max} = 0,25$. Данный факт

свидетельствует о том, что при меньшем значении σ_{max} ограничение по СКО самостоятельно фильтрует ответы и до ограничения по энтропии пограничные ответы не доходят. Однако при превышении значения 0,25 ограничения по СКО к фильтру по энтропии начинают доходить спорные решения. Т.е. при значении σ_{max} ниже

0,25 на метрики влияет ограничение по СКО, а выше – влияет ограничение по энтропии.

Тепловая карта для модифицированного Accuracy Rate ведет себя более плавно – это означает, что правило не ломается при ослаблении фильтров, что также подтверждается небольшим разбросом в 8 п.п между крайними значениями.

На тепловой карте Accuracy Rate разброс гораздо больше, и он практически идентичен покрытию, что свидетельствует о наличии взаимосвязи. Т.е. рост Accuracy Rate обеспечивается не за счет более точных ответов, а за счет того, что правило отвечает на большее число объектов.

Тепловая карта доли нерелевантных решений показывает, что расширение покрытия за пределы $\sigma_{max} = 0,25$ сопровождается резким скачком нерелевантных ответов из-за того, что правило начинает отвечать на ситуации, где модели все больше расходятся.

В совокупности графики демонстрируют ожидаемое поведение – при ослаблении ограничений растут покрытие, Accuracy Rate и доля нерелевантных решений, а модифицированный Accuracy Rate падает. И наоборот – при ужесточении ограничений наблюдается обратная картина.

Стоит сделать оговорку, эксперимент в целом показывает поведение правила «к-консенсус» при регулировании параметров, однако конкретные числа могут отличаться в зависимости от выбранного датасета.

Проведенное исследование чувствительности разработанного правила доказывает, что оно управляемое и его можно настраивать под требования задачи. При этом нет опасных значений параметров, при которых качество правила резко обваливается. При этом стоит отметить, что асимметрия по $k_{нижн}$ и $k_{верх}$ – не свойство правила, а следствие несбалансированности датасета.

VII. ЗАКЛЮЧЕНИЕ

Таким образом разработано параметризованное пороговое правило принятия решений группой искусственных интеллектуальных акторов и названо «к-консенсус», k обозначает границы области неопределенности, а слово *консенсус* подразумевает принятие решений в группе, данные части названия отражают суть разработанного правила. Экспериментальная проверка разработанного правила показала, что оно значительно выигрывает у классических правил голосования в случаях, когда допускается отказаться от решения в спорных ситуациях, т.е. его область применения – ситуации с высокой стоимостью ошибки. Отдельно проведено исследование чувствительности правила к изменению параметров, что показало его устойчивость и управляемость.

Данное правило позволяет расширить область применения метода эволюционного согласования на несоставные решения и упростить этап экспертизы решений с сохранением преимуществ данного метода. Отдельно стоит сказать, что данное правило имеет

ограничение – т.к. оно опирается на генеральную среднюю метрику уверенности акторов, то модели должны быть обучены выдавать соответствующий показатель, а значит формирование составных решений в текущем виде правила не представляется возможным.

Выведены формулы, позволяющие определить значения границ для формирования областей отрицательных, положительных решений и неопределенности. А также представлено необходимое и достаточное условие для существования области неопределенности при принятии решений группой искусственных интеллектуальных акторов.

БИБЛИОГРАФИЯ

- [1] Соколов Е.Г., Протасов В.И., Бурханов Р.А., Потапова З.Е. Использование метода эволюционного согласования при нейросетевом распознавании рукописных цифр // Труды международной научной конференции СРТ1617, Ларнака, Республика Кипр, 7-14 мая 2017 г. Протвино: АНО «Институт физико-технической информатики», 2017. С. 331-334.
- [2] Du Y., Li S., Torralba A., Tenenbaum J.B., Mordatch I. Improving Factuality and Reasoning in Language Models through Multiagent Debate // arXiv.org e-Print archive. 2023. P. 1-27. URL: <https://doi.org/10.48550/arXiv.2305.14325> (дата обращения 21.07.2026).
- [3] Качалин В.С. Применение группы искусственных интеллектуальных акторов для принятия релевантного решения о проведении косметического ремонта // Современные наукоемкие технологии. Региональное приложение. 2025. №3(83). С. 83-89.
- [4] Протасов В.И., Шарнин М.М., Потапова З.Е., Мельников Е.И., Морочко Минчала Х.П. Теорема Кондорсе и теоремы об экспертах // Труды Международной научной конференции СРТ2015, Ларнака, Республика Кипр, 10-17 мая 2015 г. Протвино: АНО «Институт физико-технической информатики», 2016. С. 277-279.
- [5] Протасов В.И. Метод эволюционного согласования решений. Компьютерная и математическая модели // Горный информационно-аналитический бюллетень (научно-технический журнал). 2011. №51. С. 360-379.
- [6] Мирахмедов Р.О. Математическая модель итерационного метода эволюционного согласования решений // International Journal of Open Information Technologies. 2024. №6. С. 82-91.
- [7] Маркарян Л.В. Применение метода эволюционного согласования решений для форсайт-проектов в промышленности // Горный информационно-аналитический бюллетень (научно-технический журнал). 2013. №7. С. 175-178.
- [8] Rain in Australia [Электронный ресурс] // kaggle. – URL: <https://www.kaggle.com/datasets/jsphyg/weather-dataset-rattle-package>.

Качалин Василий Сергеевич,
ведущий программист ИТО ПХП ФГАОУ ВО
«Московский авиационный институт (национальный
исследовательский университет)», Москва, РФ
AuthorID: 1183400, SPIN: 9391-7395, ORCID: 0000-0002-
7370-109X email: vasilij.kachalin@gmail.com

Протасов Владислав Иванович,
д.т.н, доцент, профессор каф. 307 ФГАОУ ВО
«Московский авиационный институт (национальный
исследовательский университет)», Москва, РФ
AuthorID: 520803, SPIN: 9455-1291, ORCID: 0000-0002-
4783-7209 email: protonus@yandex.ru

The «k-consensus» decision-making rule for a group of artificial intelligent actors at the second stage of the evolutionary matching method

Vasilii Kachalin, Vladislav Protasov

Abstract — The article presents a decision-making rule for a group of artificial intelligent actors called «k-consensus», designed for the second stage of the evolutionary matching method – the stage of evaluating actors' responses. The developed rule makes it possible to analyze the responses of several machine learning classification models and, based on this analysis, to form a group decision that, according to Condorcet's theorem, has an advantage over individual decisions in terms of accuracy and relevance. In addition to positive and negative decisions, the developed rule allows the group of actors to express doubt and respond with “I don't know”, which implies the transfer of the right to make a decision to a human. This response is ensured by evaluating the decisions of the actors using the root-mean-square deviation, Shannon entropy, and the general average of the confidence metrics for each actor. This rule has a limitation – it can be applied only to non-composite decisions of actors. Formulas have been derived to calculate the values of the boundaries of the group's uncertainty zone, and a necessary and sufficient condition for the existence of this zone has been presented. Experimental testing of the developed rule has shown that it significantly outperforms classical voting rules in cases where it is permissible to refrain from making a decision in disputed situations. A study of the rule's sensitivity to changes in parameters was conducted, which demonstrated its stability and controllability.

Key words — a group of actors, artificial intelligence, the evolutionary matching method, decision areas, decision-making, ternary logic, Shannon entropy.

REFERENCES

- [1] Sokolov E.G., Protasov V.I., Burhanov R.A., Potapova Z.E. Using the method of evolutionary matching for neural network recognition of handwritten figures // Proceedings of the International Scientific Conference CPT1617, Larnaca, Republic of Cyprus, 7-14 may 2017 Protvino: ANO « Institute of Physical and Technical Informatics», 2017. P. 331-334.
- [2] Du Y., Li S., Torralba A., Tenenbaum J.B., Mordatch I. Improving Factuality and Reasoning in Language Models through Multiagent Debate // arXiv.org e-Print archive. 2023. P. 1-27. URL: <https://doi.org/10.48550/arXiv.2305.14325> (date of request 21.07.2026).
- [3] Kachalin V.S. The use of a group of artificial intelligent actors to make a relevant decision on cosmetic repairs // Modern high technologies. Regional application. 2025. No. 3(83). P. 83-89.
- [4] Protasov V.I., Shamin M.M., Potapova Z.E., Mel'nikov E.I., Morocho Minchala H.P. Condorcet's jury theorem and expert theorems // Proceedings of the International Scientific Conference CPT2015, Larnaca, Republic of Cyprus, 10-17 may 2015. Protvino: ANO « Institute of Physical and Technical Informatics», 2016. P. 277-279.
- [5] Protasov V.I. The method of evolutionary solution coordination. Computer and mathematical models. // Mining informational and analytical bulletin (scientific and technical journal). 2011. No. S1. P. 360-379.
- [6] Mirakhmedov R.O. Mathematical model of the iterative method of evolutionary coordination of solutions // International Journal of Open Information Technologies. 2024. No. 6. P. 82-91.
- [7] Markaryan L.V. Evolutionary solution concordance method application in the industrial foresight projects // Mining informational and analytical bulletin (scientific and technical journal). 2013. No. 7. P. 175-178.
- [8] Rain in Australia [Electronic resource] // kaggle. – URL: <https://www.kaggle.com/datasets/jsphyg/weather-dataset-rattle-package>.

Vasilii Kachalin,

Lead software developer, Information and Technical Department of the Production and Economic Units of the Moscow Aviation Institute
AuthorID: 1183400, SPIN: 9391-7395, ORCID: 0000-0002-7370-109X email: vasilii.kachalin@gmail.com

Vladislav Protasov,

Associate Professor, Doctor of Technical Sciences, Moscow Aviation Institute, Professor of the Department of Digital Technologies and Information Systems
AuthorID: 520803, SPIN: 9455-1291, ORCID: 0000-0002-4783-7209 email: protonus@yandex.ru