

ГИБРИДНАЯ МОДЕЛЬ И АЛГОРИТМ ИНТЕЛЛЕКТУАЛЬНОЙ ОЦЕНКИ ЭФФЕКТИВНОСТИ ТРУДА СОТРУДНИКОВ НА ОСНОВЕ МЕТОДОВ KNN И RANDOM FOREST

Зокиров Санжар Икромжон угли

д-р филос. физ.-мат. наук (PhD), доцент,

Ферганский государственный технический университет, Узбекистан, г. Фергана

Каримова Гульшаной Акмалжон кизи

Ферганский государственный технический университет, Узбекистан, г. Фергана

Аннотация. Статья посвящена разработке модели и алгоритмов интеллектуальной оценки эффективности труда сотрудников на основе гибридных методов машинного обучения. Показано, что существующие подходы к оценке персонала — от традиционных бумажных форм до современных электронных HR-систем — опираются на фиксированный набор критериев и не используют методы машинного обучения для классификации уровня эффективности, что ограничивает их точность и адаптивность. Предложена математическая модель, представляющая эффективность сотрудника в виде нормализованного многомерного вектора признаков по пяти измерениям деятельности, а также модель адаптивного формирования и взвешивания критериев на основе метода анализа иерархий (МАИ/АНР) и корреляционного анализа Пирсона. Разработан гибридный алгоритм классификации, объединяющий метод k-ближайших соседей (KNN) и случайный лес (Random Forest) через механизм маршрутизации по уровню достоверности с калиброванным порогом. Порог отбирался методом пятикратной перекрёстной проверки. Экспериментальная проверка показала повышение точности классификации с 76,2% до 92,8% и сокращение времени обработки данных с 12,4 до 6,1 секунды по сравнению с использованием одного базового алгоритма. Полученные результаты подтверждают эффективность предложенного гибридного подхода и могут быть использованы при построении интеллектуальных систем управления персоналом.

Ключевые слова: оценка эффективности труда, гибридные алгоритмы машинного обучения, метод k-ближайших соседей, случайный лес, метод анализа иерархий, многокритериальный анализ, управление персоналом.

HYBRID MODEL AND ALGORITHM FOR INTELLIGENT ASSESSMENT OF EMPLOYEE PERFORMANCE BASED ON KNN AND RANDOM FOREST METHODS

Zokirov Sanjar Ikromjon ugli

PhD (Physics and Mathematics), Associate Professor,
Fergana State Technical University, Fergana, Uzbekistan

Karimova Gulshanoy Akmaljon kizi

Fergana State Technical University, Fergana, Uzbekistan

Abstract. This article focuses on the development of a model and algorithms for the intelligent assessment of employee performance using hybrid machine learning methods. It demonstrates that existing personnel assessment approaches—ranging from traditional paper-based forms to modern electronic HR systems—rely on a fixed set of

criteria and fail to utilize machine learning methods to classify performance levels, thereby limiting their accuracy and adaptability. A mathematical model is proposed that represents employee performance as a normalized multidimensional feature vector across five activity dimensions, alongside a model for the adaptive generation and weighting of criteria based on the Analytic Hierarchy Process (AHP) and Pearson correlation analysis. A hybrid classification algorithm was developed, combining the k-Nearest Neighbors (KNN) method and Random Forest via a confidence-level routing mechanism with a calibrated threshold. The threshold was selected using five-fold cross-validation. Experimental testing demonstrated an increase in classification accuracy from 76.2% to 92.8% and a reduction in data processing time from 12.4 to 6.1 seconds compared to using a single base algorithm. The obtained results confirm the effectiveness of the proposed hybrid approach and can be used in the development of intelligent personnel management systems.

Keywords: performance assessment, hybrid machine learning algorithms, k-nearest neighbors method, random forest, Analytic Hierarchy Process, multi-criteria analysis, personnel management.

Введение

Повышение объективности и оперативности оценки эффективности труда сотрудников является одной из ключевых задач современного управления персоналом. В условиях роста масштабов организаций и усложнения структуры трудовой деятельности традиционные методы оценки — периодические аттестации, экспертные заключения руководителей, бумажные формы отчётности — не обеспечивают необходимой скорости обработки данных и не позволяют учитывать множество разнородных критериев одновременно [1].

Анализ существующих решений показывает, что современные подходы к оценке персонала можно условно разделить на три группы: традиционные методы (периодические письменные отчёты и аттестации), методы целевого управления (Management by Objectives, система ключевых показателей эффективности KPI, метод «360 градусов») и электронные системы управления персоналом (HRIS, e-HRM). Несмотря на автоматизацию отдельных процессов, электронные системы по-прежнему используют жёстко заданные, статические наборы критериев оценки, изменение которых требует модификации структуры базы данных [2].

На основании проведённого анализа выделены три взаимосвязанные проблемы, определяющие научную задачу исследования. Во-первых, критерии оценки жёстко привязаны к структуре базы данных, что затрудняет их адаптацию под специфику конкретной организации. Во-вторых, существующие алгоритмы классификации сталкиваются с компромиссом между скоростью и точностью: быстрые методы, такие как KNN, уступают в точности на пограничных случаях, тогда как более точные ансамблевые методы, такие как Random Forest, требуют значительных вычислительных ресурсов [3, 4]. В-третьих, отсутствуют оптимизированные модели, специально предназначенные для многокритериальной, управляемой данными оценки эффективности труда.

Целью настоящего исследования является разработка моделей и алгоритмов интеллектуальной оценки эффективности труда сотрудников на основе гибридных методов машинного обучения, обеспечивающих повышение точности оценки и вычислительной эффективности процесса классификации.

Для достижения поставленной цели решались следующие задачи: 1) разработка формализованной многокритериальной математической модели эффективности труда сотрудника; 2) разработка модели адаптивного

формирования и взвешивания критериев оценки на основе метода анализа иерархий и корреляционного анализа; 3) разработка гибридного алгоритма классификации, объединяющего KNN и Random Forest с маршрутизацией по уровню достоверности, а также метода выбора оптимального порогового значения; 4) создание программного комплекса и экспериментальная проверка предложенных моделей и алгоритмов.

Материалы и методы

Формальная модель эффективности. Эффективность труда сотрудника формализована в виде нормализованного вектора признаков $x = (x_1, x_2, x_3, x_4, x_5)$, где компоненты соответствуют пяти измерениям трудовой деятельности: выполнение задач, качество работы, профессиональное поведение, взаимодействие и коммуникация, развитие и инициативность. Каждый компонент приводится к диапазону $[0, 1]$ методом min-max нормализации:

$$x_i' = \frac{x_i - x_{i \min}}{x_{i \max} - x_{i \min}} \quad (1)$$

что обеспечивает сопоставимость разнородных по природе показателей и корректность последующего вычисления расстояний между наблюдениями.

Модель формирования критериев. Для формирования системы весовых коэффициентов критериев применён метод анализа иерархий (МАИ, АНР): на основе матрицы парных сравнений экспертов вычисляются относительные веса критериев w_i , при этом согласованность суждений контролируется индексом согласованности (Consistency Ratio) [5]. Для устранения избыточности показателей дополнительно проводится корреляционный анализ Пирсона: критерии с высоким коэффициентом взаимной корреляции ($r > 0,85$) объединяются либо исключаются из модели, что снижает мультиколлинеарность признакового пространства.

Гибридный алгоритм классификации. Классификация уровня эффективности сотрудника выполняется гибридным алгоритмом, объединяющим метод k-ближайших соседей и случайный лес через механизм маршрутизации по уровню достоверности. На первом этапе алгоритм KNN определяет класс объекта на основе взвешенного евклидова расстояния до k ближайших соседей в обучающей выборке и одновременно формирует оценку достоверности классификации $s \in [0, 1]$, отражающую степень согласованности соседей:

$$d(x, x_j) = \sqrt{\sum w_i (x_i - x_{ji})^2}, \quad (2)$$

Если оценка достоверности превышает пороговое значение θ , результат KNN принимается как окончательный (быстрый путь обработки). В противном случае, когда $s \leq \theta$, наблюдение считается пограничным и передаётся на дообработку алгоритму Random Forest, ансамблю решающих деревьев, обеспечивающему более устойчивую классификацию за счёт агрегирования решений [3]. Такой подход позволяет сочетать вычислительную скорость KNN с точностью ансамблевого метода без потери качества классификации.

Оптимальное пороговое значение θ определялось методом пятикратной перекрёстной проверки (5-fold cross-validation) на диапазоне значений от 0,5 до 0,95 с шагом 0,02. Критерием выбора служило сочетание максимальной точности классификации и минимальной доли наблюдений, направляемых на дообработку алгоритмом Random Forest. По результатам перекрёстной проверки оптимальным признано значение $\theta = 0,72$, при котором около 78% наблюдений обрабатываются по быстрому пути.

Программная реализация. Для верификации предложенных моделей разработан программный комплекс, включающий реляционную базу данных из

десяти сущностей в третьей нормальной форме, модуль формирования и хранения критериев оценки, реализацию гибридного алгоритма классификации и модуль формирования рекомендаций по ключевым показателям эффективности. Архитектура комплекса допускает реконфигурацию критериев оценки без изменения структуры базы данных, что соответствует задаче адаптивного формирования критериев.

Результаты

Экспериментальная проверка предложенного гибридного алгоритма проводилась в сравнении с применением одного базового алгоритма классификации (без маршрутизации по достоверности). Сравнительные результаты по основным метрикам качества классификации представлены в таблице 1.

Точность (Accuracy)	76,2%	92,8%	+16,6 п.п.
Точность (Precision)	75,1%	91,3%	+16,2 п.п.
Полнота (Recall)	74,3%	90,5%	+16,2 п.п.
F1-мера	74,7%	90,9%	+16,2 п.п.
Время обработки	12,4 с	6,1 с	-50,8%

Таблица 1. Сравнение метрик качества классификации до и после применения гибридного алгоритма

Как видно из таблицы 1, применение гибридного алгоритма обеспечило прирост точности классификации на 16,6 процентных пункта при одновременном сокращении времени обработки более чем в два раза. Показатели точности (precision), полноты (recall) и F1-меры превысили 90%, что свидетельствует об отсутствии значимого смещения классификатора в сторону какого-либо из классов эффективности.

Анализ зависимости точности классификации от порогового значения θ показал немонотонный характер: при заниженных значениях порога ($\theta < 0,6$) на дообработку Random Forest направляется избыточно большая доля наблюдений, что увеличивает время обработки без существенного прироста точности; при завышенных значениях ($\theta \geq 0,9$) точность классификации снижается до 87,4%, поскольку часть пограничных случаев ошибочно принимается по результатам KNN без дополнительной проверки. Значение $\theta = 0,72$ обеспечило оптимальное соотношение между точностью и вычислительными затратами.

Разработанный программный комплекс подтвердил практическую применимость предложенных моделей: реализованная база данных из десяти сущностей позволяет хранить многомерные данные об эффективности сотрудников без потери гибкости при изменении состава критериев, а модуль формирования критериев на основе МАИ и корреляционного анализа обеспечивает исключение избыточных показателей на этапе настройки системы под конкретную организацию.

Обсуждение

Полученные результаты подтверждают гипотезу о том, что объединение быстрого, но менее устойчивого к пограничным случаям алгоритма KNN с более

точным, но вычислительно затратным алгоритмом Random Forest через механизм маршрутизации по достоверности позволяет получить решение, превосходящее по совокупности показателей каждый из базовых алгоритмов в отдельности. Аналогичный вывод о преимуществе гибридных и ансамблевых схем над одиночными классификаторами при решении задач классификации с несбалансированными или неоднородными признаками отмечается и в более ранних работах [3, 4].

Предложенная модель формирования критериев на основе МАИ и корреляционного анализа отличается от подходов, используемых в существующих HR-системах, тем, что позволяет адаптировать состав и веса критериев под специфику организации без изменения структуры базы данных, тогда как в системах типа HRIS и e-HRM набор критериев остаётся статическим и требует программной доработки при изменении требований к оценке [2].

Вместе с тем необходимо отметить ряд ограничений выполненного исследования. Во-первых, найденное пороговое значение $\theta = 0,72$ является оптимальным для использованного набора данных и организационного контекста; при переносе модели на данные другой организации требуется повторная калибровка порога методом перекрёстной проверки. Во-вторых, для окончательной оценки эффективности предложенного гибридного алгоритма необходимо его сопоставление с альтернативными ансамблевыми методами (например, градиентным бустингом) на выборках большего объёма, что определяет направление дальнейших исследований. В-третьих, оценка вычислительной сложности алгоритма выполнена эмпирически и подлежит строгому теоретическому обоснованию.

Дальнейшее развитие исследования предполагает: теоретическое обоснование и формализацию предложенной многокритериальной модели; сравнительный анализ гибридного алгоритма с иными ансамблевыми методами классификации на данных большего объёма и из разных организаций; а также практическое внедрение разработанного программного комплекса в деятельность организаций с последующей оценкой экономического эффекта.

Заключение

В работе предложены математическая модель многокритериальной оценки эффективности труда сотрудников, модель адаптивного формирования и взвешивания критериев на основе метода анализа иерархий и корреляционного анализа, а также гибридный алгоритм классификации, объединяющий методы KNN и Random Forest посредством маршрутизации по уровню достоверности с калиброванным пороговым значением $\theta = 0,72$. Экспериментальная проверка на программном комплексе показала повышение точности классификации с 76,2% до 92,8% и сокращение времени обработки данных с 12,4 до 6,1 секунды по сравнению с использованием одного базового алгоритма.

Полученные результаты имеют как теоретическое значение — предложенная схема маршрутизации по достоверности может быть распространена на другие задачи многокритериальной классификации, — так и практическую значимость для построения интеллектуальных систем управления персоналом, обеспечивающих гибкую настройку критериев оценки без изменения структуры информационной системы. Дальнейшие исследования будут направлены на теоретическое обоснование модели, сопоставление с альтернативными ансамблевыми методами и практическое внедрение результатов в организациях различного профиля.

Литература:

1. Kiron, D., Schrage, M., Candelon, F., Khodabandeh, S., & Chu, M. (2023). Strategic alignment with AI and smart KPIs. *MIT Sloan Management Review*, 65(1), 1-6.
2. Karimova G. System for determining and evaluating labour effectiveness criteria. State software registration certificate DGU №53438, 2025.
3. Зокиров, С. И. У., & Каримова, Г. А. К. (2025). Разработка системы формирования критериев для определения эффективности труда сотрудников (KPI) и методы вычисления оценки. *Universum: технические науки*, 1(11 (140)), 40-44.
4. Елкина, О.С.(2026). Риски искусственного интеллекта для менеджмента: классификация, управление, KPI. *Вестник Московского городского педагогического университета. Серия: Экономика*, (1 (47)), 163-187.
5. Armstrong M. *Armstrong's Handbook of Performance Management: An Evidence-Based Guide to Delivering High Performance*. 6th ed. London: Kogan Page, 2017. 288 p.
6. DeNisi A. S., Murphy K. R. Performance appraisal and performance management: 100 years of progress? // *Journal of Applied Psychology*. 2017. Vol. 102, № 3. P. 421–433.
7. Breiman L. Random Forests // *Machine Learning*. 2001. Vol. 45, № 1. P. 5–32.
8. Saaty T. L. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. New York: McGraw-Hill, 1980. 287 p.
9. Hastie T., Tibshirani R., Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York: Springer, 2009. 745 p.