УДК 004.89
DOI: 10.26102/2310-6018/2025.49.2.014

Повышение достоверности объяснимого искусственного интеллекта посредством нечеткой логики и онтологии

П.И. Косов✉, Л.А. Гардашова

Азербайджанский государственный университет нефти и промышленности, Баку, Азербайджан

Резюме. Недостаточная объяснимость моделей машинного обучения длительное время являлась существенной проблемой. Специалисты в различных областях применения искусственного интеллекта (ИИ) стремились к созданию объяснимых и надежных систем. Для решения данной проблемы DARPA разработала современный подход к объяснимому ИИ (XAI). Впоследствии Bellucci и др. расширили концепцию XAI от DARPA, предложив новый метод, основанный на технологиях семантической паутины. В частности, они использовали онтологии OWL2 для представления экспертных знаний, ориентированных на пользователя. Данная система повышает доверие к решениям ИИ путем предоставления более глубоких объяснений. Тем не менее, системы XAI по-прежнему испытывают затруднения в условиях неполных и неточных данных. Мы предлагаем новый подход, использующий нечеткую логику для решения этой проблемы. Наша методика основана на сочетании нечеткой логики и моделей машинного обучения для имитации человеческого мышления. Данный новый подход более эффективно взаимодействует с экспертными знаниями для обеспечения более глубоких объяснений решений ИИ. Система использует экспертные знания, представленные в виде онтологий, что полностью соответствует архитектуре, предложенной Bellucci и др. в их работе. Целью данной работы является не улучшение точности классификации данных, а повышение достоверности и глубины объяснений, полученных от XAI с использованием «объяснимых» свойств и нечёткой логики.

Ключевые слова: объяснимый искусственный интеллект, объяснимость, онтология, нечеткая система, нечеткая кластеризация.

Для цитирования: Косов П.И., Гардашова Л.А. Повышение достоверности объяснимого искусственного интеллекта посредством нечеткой логики и онтологии. *Моделирование, оптимизация и информационные технологии.* 2025;13(2). URL: https://moitvivt.ru/ru/journal/pdf?id=1872 DOI: 10.26102/2310-6018/2025.49.2.014

Enhancing the trustworthiness of explainable artificial intelligence through fuzzy logic and ontology

P.I. Kosov✉, L.A. Gardashova

Azerbaijan State Oil and Industry University, Baku, Azerbaijan

Abstract. The insufficient explainability of machine learning models has long constituted a significant challenge in the field. Specialists across various domains of artificial intelligence (AI) application have endeavored to develop explicable and reliable systems. To address this challenge, DARPA formulated a contemporary approach to explainable AI (XAI). Subsequently, Bellucci et al. expanded DARPA's XAI concept by proposing a novel methodology predicated on semantic web technologies. Specifically, they employed OWL2 ontologies for the representation of user-oriented expert knowledge. This system enhances confidence in AI decisions through the provision of more profound explanations. Nevertheless, XAI systems continue to encounter difficulties when confronted with incomplete and imprecise data. We propose a novel approach that utilizes fuzzy logic to address this limitation. Our methodology is founded on the integration of fuzzy logic and machine learning models to imitate human thinking. This new approach more effectively interfaces with expert knowledge to facilitate deeper

© Косов П.И., Гардашова Л.А., 2025

explanations of AI decisions. The system leverages expert knowledge represented through ontologies, maintaining full compatibility with the architecture proposed by Bellucci et al. in their work. The objective of this research is not to enhance classification accuracy, but rather to improve the trustworthiness and depth of explanations generated by XAI through the application of "explanatory" properties and fuzzy logic.

Keywords: explainable artificial intelligence, explainability, ontology, fuzzy system, fuzzy clustering.

For citation: Kosov P.I., Gardashova L.A. Enhancing the trustworthiness of explainable artificial intelligence through fuzzy logic and ontology. *Modeling, Optimization and Information Technology*. 2025;13(2). (In Russ.). URL: https://moitvivt.ru/ru/journal/pdf?id=1872 DOI: 10.26102/2310-6018/2025.49.2.014

Введение

Потребность экспертов и пользователей в прозрачности выходных данных систем искусственного интеллекта обусловила разработку объяснимых систем. Область искусственного интеллекта (ИИ, англ. Artificial Intelligence, AI), способная предоставлять объяснения результатов работы моделей машинного обучения, получила название объяснимого ИИ (англ. eXplainable AI, XAI) [1]. Алгоритмы машинного обучения рассматриваются как основной метод для разработки ИИ [2]. Они базируются на статистических алгоритмах и способны выполнять задачи без явных команд. Можно утверждать, что XAI уже нашел применение в широком спектре предметных областей [3]. Мы полагаем, что повышение объяснимости машинного обучения и обучающихся моделей может способствовать повышению доверия к принимаемым ими решениям. Основным направлением этой статьи является не улучшение точности алгоритма кластеризации и классификации, а показать пример, как возможно улучшить объяснения системы XAI при помощи нечеткого подхода.

Gunning и др. [4] рассматривают систему XAI, разработанную DARPA, которая объединяет модели машинного обучения и интерфейс объяснений в единую архитектуру. Данный подход обеспечивает четкое представление о том, как модели принимают определенные решения. Тем не менее, для повышения объяснимости Bellucci и др. [5] предлагают усовершенствование архитектуры DARPA путем интеграции онтологического представления экспертных знаний. Новая система классификатор изображений на основе онтологии (англ. Ontology-based Image Classifier, OBIC) полностью согласуется с подходом DARPA [4] с учетом дополнительных модификаций.

Объединение онтологического представления предметных знаний с методами машинного обучения широко применяется в большинстве систем объяснимого искусственного интеллекта (XAI). Kulmanov и др. [6] подробно рассмотрели, как сочетание машинного обучения и онтологий помогает обнаруживать семантические сходства в данных и показывать, как онтологии ограничивают и обогащают модели. Giustozzi и др. [7] представили онтологический подход в Индустрии 4.0 для совершенствования моделирования данных от устройств, модулей и датчиков, интегрируя эти данные с предметными знаниями и применяя потоковые рассуждения для интеграции данных из множества источников в реальном времени. Bourgais и др. [8] продемонстрировали необходимость семантических систем в интернете вещей и здравоохранении, где онтологии хранят данные, получаемые с различных датчиков, и применяют потоковые рассуждения для выявления проблем со здоровьем.

Для улучшения точности алгоритмы машинного обучения могут быть интегрированы с нечеткой логикой [9] в рамках «мягких вычислений» (англ. Soft Computing). Aliev R.A. и др. в своей работе [10] представили основы теории мягких вычислений, описывая различные приложения теории нечётких множеств, нейронных

сетей, генетических алгоритмов, теории хаоса и других. Dumitrescu и др. [11] подчеркнули практические применения нечеткой логики в улучшении автоматизированного интеллектуального управления в динамических и неопределенных средах. Также, Gardashova L.A. [12] представила математический подход к решению задач оптимального управления с помощью нечеткой логики, демонстрируя применение нечеткого реляционного уравнения для нечетких состояний и систем управления.

Цель исследования заключается в том, чтобы улучшить доверие к результатам объяснимой системы. Основной фокус текущего исследования направлен на представление экспертных знаний, а также улучшение достоверности объяснения, не рассматривая точность алгоритма классификации. Мы рассматриваем применение нашего нового метода, использующего нечеткую логику [9]. Предложенный подход основан на архитектуре OBIC [5], а также используются «объяснительные» свойства (англ. "explanatory" properties) [13]. Согласно нашим наблюдениям, неполные и неточные данные не позволяют XAI системам обеспечивать полностью прозрачный вывод. Мы рассматриваем надежный XAI как систему, которая не только предоставляет объяснение для результатов классификации, но и справляется с неопределенностью наборов данных реального мира. В результате нами предложена новая система, которая интегрирует нечеткую логику с объяснительными свойствами в онтологии, стремясь повысить интерпретируемость и надежность систем XAI, сохраняя при этом их способность справляться со сложностью и неопределенностью информации и знаний из реального мира. Улучшение точности классификации не является целью данной работы.

Материалы и методы

В данном разделе мы подробно рассматриваем все методы, использованные в нашей работе. Была исследована новая система, объединяющая несколько передовых подходов. Мы использовали «объяснительные» свойства [13] в архитектуре современной объяснимой системы на основе онтологий (OBIC) [5] с применение нечеткого подхода на основе метода нечеткой кластеризации C-средних [14].

Краткое определение онтологии. Литература по онтологиям предоставляет различные и иногда противоречивые определения. Онтологии предоставляют определения для объектов, их экземпляров и отношений между ними, наряду с аксиомами. Это вид структурированных знаний, который высоко применим для улучшения интерпретируемости и прозрачности модели ИИ. Онтология наделяет систему XAI способностью предоставлять экспертные знания из внешнего мира и создавать гораздо более точные и значимые объяснения для преодоления разрыва между необработанными данными и осмысленными выводами. В формуле (1) показано, как онтология может быть определена:

$$O = (C, \leq_C, R, \leq_R, A^O). \quad (1)$$

В уравнении (1) сама онтология представлена как O. Множество концепций объектов представлены как C и имеют частичный порядок $\leq_C$. R является набором отношений между объектами, который частично упорядочен $\leq_R$, и представлено как $R \subseteq C \times C$. Также A^O является множеством аксиом для онтологии O. В случае (a, b) принадлежит R, это может быть записано как $b = r(a)$.

Классификатор изображений на основе онтологий (OBIC). OBIC [5] имеет два основных компонента: первый компонент – это множественные модели машинного обучения для классификации свойств и главного класса, и второй компонент – это интеграция семантических веб-технологий для обеспечения предоставления экспертных знаний. Система ограничена входными данными, которые должны быть только

изображениями. Ключевая идея OBIC опирается на визуальные свойства, которые находятся в объектах на изображениях и описываются в онтологии экспертами. Для правильной обработки данных эксперт вручную создает онтологию. Архитектура OBIC взята из диссертации[1] Bellucci на 86 странице и представлена на Рисунке 1.

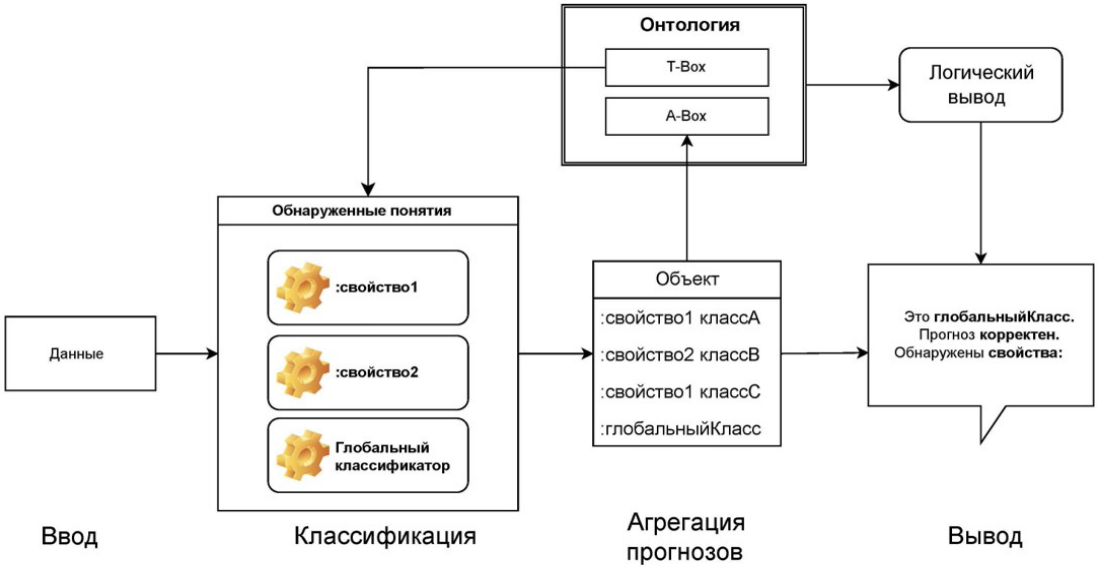

Рисунок 1 – Общая схема функционирования OBIC
Figure 1 – General scheme of OBIC functioning

OBIC использует OWL2 Web Ontology Language[2] и применяет предположение открытого мира (англ. Open World Assumption, OWA). OWA предполагает, что отсутствие объектов некорректно, что означает, что утверждение остается истинным, пока не доказано обратное. Онтология должна точно отражать входные данные, четко включая как положительные, так и отрицательные ограничения, чтобы обеспечить достоверность и ясность в объяснениях. Тем не менее, автор утверждает, что в некоторых случаях строгая точность может привести к несогласованностям.

Подход системы OBIC заключается в том, что каждое свойство, которое должно быть найдено в объекте на изображении, связано с отдельной моделью машинного обучения. Также отдельный глобальный классификатор ответственен за идентификацию основного класса объекта. После идентификации всех классов и свойств, они организуются в онтологические индивиды (экземпляры) и включаются в ABox онтологии для дальнейшего логического рассуждения.

«Объяснительные» свойства. Эти свойства представлены в работе [13], они соответствуют всем требованиям и архитектуре OBIC, а также позволяют иметь объяснения для множественных типов данных, следуя первоначальной идее, представленной DARPA [4]. Как упоминалось ранее, система XAI не имеет объяснений, основанных на онтологиях. Возможность применения текущих свойств к табличным типам данных была обсуждена в работе [15]. Однако, OBIC фокусируется на использовании свойств в онтологии для генерации объяснений на основе визуальных характеристик данных.

[1] Bellucci M. Symbolic Approaches for Explainable Artificial Intelligence. Caen: Normandie Université; 2023. 174 p. URL: https://theses.hal.science/tel-04469103 (дата обращения: 21.02.2025).
[2] W3C OWL Working Group. OWL 2 Web Ontology Language Document Overview (Second Edition). World Wide Web Consortium. URL: https://www.w3.org/TR/owl2-overview (дата обращения: 21.02.2025).

В работе [16] предлагается использовать логическое рассуждение и ментальные модели для «объяснительных» свойств, которые адаптированы для соответствия уровня экспертизы пользователя и предыдущим базовым знаниям. Авторы считают, что соответствующие ментальные модели должны использоваться для обоснования каждого объяснения, учитывая уровень знаний и экспертизу пользователя. Независимо от типа данных, наличие «объяснительного» атрибута делает XAI более точным в своих объяснениях [15]. Это облегчает пользователям идентификацию и исправление системных ошибок.

Нечеткая кластеризация С-средних (англ. Fuzzy C-means, FCM). Данный алгоритм применяется в среде неопределенных данных, когда каждый входной параметр может принадлежать нескольким различным классам [14]. Объекты одновременно могут быть членами других кластеров или классов. Им присваиваются степени принадлежности для каждого класса, и значение варьируется в диапазоне от 0 до 1.

Существуют два важных уравнения для вычисления степеней принадлежности (2) и их обновления (3), когда это необходимо. Формула (2) учитывает отношение точки к каждому кластеру, определяя степень соответствия каждой точки с центрами этих кластеров.

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{||x_i - c_j||}{||x_i - c_k||}\right)^{\frac{2}{m-1}}}, \quad (2)$$

где μ_{ij} – это вычисленное значение, представляющее степень принадлежности точки данных i к кластеру j. $||x_i - c_j||$ и $||x_i - c_k||$ вычисляют евклидовы расстояния между точкой данных и кластером. Чем меньше расстояние, тем выше уровень принадлежности к конкретному кластеру. Их отношение дает каждой точке данных возможность быть частично или полностью связанной с несколькими кластерами путем оценки степеней принадлежности. Количество нечеткости кластера контролируется параметром нечеткости m.

Другое уравнение (3) используется для обновления значения степени принадлежности при вычислении точек данных согласно новому кластеру.

$$c_j = \frac{\sum_{i=1}^{n}(\mu_{ij})^m \cdot x_i}{\sum_{i=1}^{n}(\mu_{ij})^m}, \quad (3)$$

где c_j – новая позиция для кластера j; центроид кластера. Также x_i – это точка данных, а $(\mu_{ij})^m$ представляет степень принадлежности этой точки данных. В результате центр кластера будет притягиваться к местоположениям, более близким к центру кластера. $\sum_{i=1}^{n}(\mu_{ij})^m \cdot x_i$ отражает сумму всех точек данных, взвешенных по степени принадлежности к кластеру j.

Обобщенный алгоритм для кластеризации FCM можно увидеть в его простой версии, предоставленной ниже.

1. Определить количество кластеров k и параметр нечеткости m.
2. Установить центроиды кластеров.
3. Определить степени принадлежности точек для каждого кластера, применяя формулу (2).
4. Обновить центроиды кластеров согласно новым кластерам, используя формулу (3).
5. Повторить шаги 3 и 4, пока не перестанут происходить изменения в значениях.

Результаты

Была разработана усовершенствованная система на базе FCM [14], с адаптацией к архитектуре OBIC [5]. Работа осуществлена с использованием «объяснительных» свойств [13]. Для логического вывода в онтологии применялись правила, разработанные экспертами на основе Языка Правил Семантической Паутины (англ. Semantic Web Rule Language, SWRL) [17]. Также, тестирование проводилось на наборе клинических данных о сердечной недостаточности[3], полностью удовлетворяющем требованиям для демонстрации глубины и достоверности объяснений, разработанной нами XAI системы.

В наборе данных каждая строчка записи является индивидуальным пациентом, который сохранен в онтологии как экземпляр класса Patient. Каждый экземпляр получен из выходных данных моделей машинного обучения и описывается с использованием существующих «объяснительных» свойств и их ограничений на основе классов онтологии. В итоге были получены 3 основных класса: пациент (Patient), целевой класс – случаи смерти пациента (DeathEvent) и атрибуты целевого для описания целевого класса (Feature).

Класс DeathEvent описывается при помощи атрибутов, построенных в онтологии в классе Feature. Изначально в первичном наборе данных присутствует 12 характеристик, описывающих причины смерти от сердечного приступа, но в итоге для облегчения работы на основе экспертных знаний было выбрано 7 характеристик (Feature) и построены классы: Diabetes, CreatininePhosphokinase, HighBloodPreassure, Smoking, Age, Platelets, Anaemia.

Класс DeathEvent принимает бинарные значения при помощи подклассов Died и Survived. Также имеются подклассы у некоторых подклассов класса Feature. Подклассы Anaemia, Diabetes, HighBloodPreassure, Smoking не имеют своих подклассов и представлены в том виде, какими являются в наборе данных из-за того, что эти атрибуты целевого класса изначально являются бинарными. Числовые, изначально не бинарные, значения для других подклассов Age, CreatininePhosphokinase, Platelets были обработаны и преобразованы в набор интервалов. Онтология отображена на Рисунке 2.

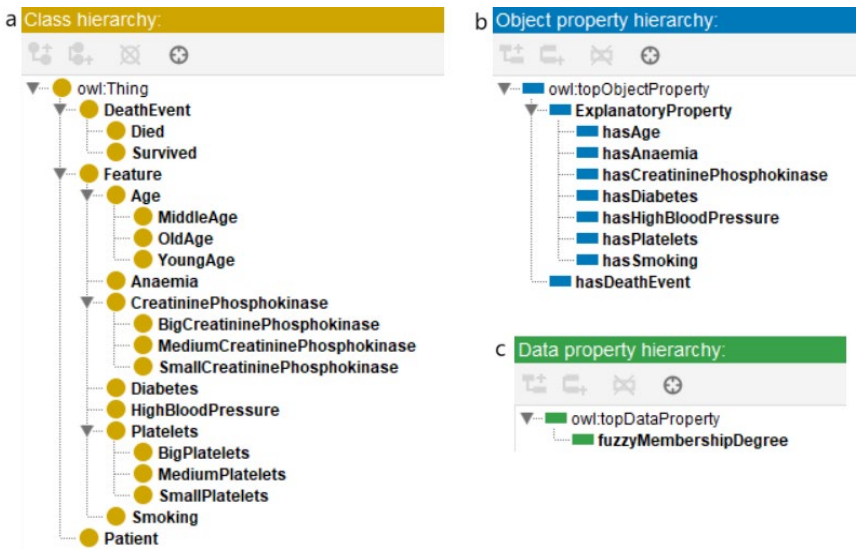

Рисунок 2 – Иерархии классов (a), свойств объектов (b) и свойств данных (c) в онтологии
Figure 2 – Hierarchies of classes (a), object properties (b), and data properties (c) in the ontology

[3] Heart Failure Clinical Records. UCI Machine Learning Repository. URL: https://doi.org/10.24432/C5Z89R (дата обращения: 06.03.2025).

Как было сказано ранее, «объяснительные» свойства основаны на экспертных знаниях и адаптированы к ментальным моделям пользователей. Таблица 1 показывает, что диапазоны (Range) и области (Domain) являются определяющими характеристиками каждого свойства.

Таблица 1 – Диапазоны (Range) и области (Domain) для всех свойств онтологии
Table 1 – Ranges and Domains for all properties of the ontology

Имена свойств	Тип свойства	Range	Domain
hasAge	«Объяснительно»	Age	Patient
hasAnaemia	«Объяснительно»	Anaemia	Patient
hasCreatininePhosphokinase	«Объяснительно»	CreatininePhosphokinase	Patient
hasDiabetes	«Объяснительно»	Diabetes	Patient
hasHighBloodPreassure	«Объяснительно»	HighBloodPreassure	Patient
hasPlatelets	«Объяснительно»	Platelets	Patient
hasSmoking	«Объяснительно»	Smoking	Patient
hasDeathEvent	Другой Object Property	DeathEvent	Patient
fuzzyMembershipDegree	Data Property	DeathEvent	xsd:decimal

Один и тот же Domain определяет свойства, в то время как Range варьируется от свойства к свойству. Причина, по которой все «объяснительные» свойства [13] имеют общий Domain, заключается в том, что все они служат для описания только экземпляров класса пациента (Patient), использующих соответствующие подклассы классовых признаков Feature в качестве своего Range. Все свойства кроме hasDeathEvent и fuzzyMembershipDegree являются «объяснительными».

«Объяснительные» свойства описывают выжившего или умершего пациента (Patient), когда hasDeathEvent определяет его класс Survived (выжил) или Died (умер). Также степень принадлежности, полученная в результате работы алгоритма нечеткой кластеризации C-средних [14], сохраняется в виде свойства данных (Data Property) fuzzyMembershipDegree. Данная степень принадлежности является показателем того, насколько класс Patient (пациент) принадлежит одному из подклассов DeathEvent.

Логический вывод выполняется с применением правил SWRL [17], разработанных экспертами. Правила используют все существующие в онтологии «объяснительные» свойства для обеспечения полной объяснимости. SWRL правила должны создаваться с учетом практической применимости в реальной жизни, обладать содержательностью и универсальностью, исключая предвзятость к определенному набору данных, чтобы обеспечить их применение к различным массивам информации, обладающей схожими свойствами с разработанной нами онтологией. На Рисунке 3 представлен пример возможных SWRL правил.

```
Patient(?p) ∧
        hasAge(?p, OldAge) ∧
        hasAnaemia(?p, true) ∧
        hasCreatininePhosphokinase(?p, MediumCreatininePhosphokinase) ∧
        hasDiabetes(?p, true) ∧
        hasHighBloodPressure(?p, true) ∧
        hasPlatelets(?p, SmallPlatelets) ∧
        hasSmoking(?p, true)
→ hasDeathEvent(?p, Died)
```

```
Patient(?p) ∧
        hasAge(?p, MiddleAge) ∧
        hasAnaemia(?p, false) ∧
        hasCreatininePhosphokinase(?p, SmallCreatininePhosphokinase) ∧
        hasDiabetes(?p, false) ∧
        hasHighBloodPressure(?p, false) ∧
        hasPlatelets(?p, MediumPlatelets) ∧
        hasSmoking(?p, false)
→ hasDeathEvent(?p, Survived)
```

Рисунок 3 – Пара возможных SWRL правил для логического вывода
Figure 3 – A pair of possible SWRL rules for reasoning

Как результат, полученные объяснения несут в себе информацию о целевом классе, также все «объяснительные» свойства [13] с их значениями. Дополнительно присутствует нечеткая степень принадлежности, которая получена от алгоритма FCM. Пример одного из объяснений от нашей системы XAI, следует ниже:

"Entry 0 for a class Patient is a Class Survived of DeathEvent. It is consistent in the ontology. Because it has the following "explanatory" properties: MiddleAge for hasAge, false for hasAnaemia, SmallCreatininePhosphokinase for hasCreatininePhosphokinase, false for hasDiabetes, false for hasHighBloodPreassure, MediumPlatelets for hasPlatelets, false for hasSmoking. The target class prediction is survived for hasDeathEvent, and it has fuzzyMembershipDegree as 0.8".

Обсуждение

Нечеткий подход помогает не просто получать бинарные результаты как «0» или «1», но благодаря нечеткой функции принадлежности формируются выходные значения в диапазоне от «0» до «1». Это помогает пользователям объяснимой системы получать намного более детальные результаты и благодаря этому делать более глубокие логические выводы. Данный метод используется чтобы предоставить пользователям достаточно глубокие объяснения и результаты с учетом всей нечеткости набора данных из реальных условий. Основное внимание уделяется исключительно объяснительным аспектам в разработанной объяснительной системе.

Однако не только нечеткий подход способствовал глубоким, достоверным и точным объяснениям. Благодаря «объяснимым» свойствам пользователь получает результат, который специально адаптирован к уровню знаний и области деятельности текущего пользователя. «Объяснимые» свойства являются ключевым аспектом вокруг которых построена система. Благодаря интеграции этих свойств и нечеткой кластеризации даются глубокие объяснения от объяснимой системы. Разработанная система является гибкой и настраивается в соответствии с потребностями конечных пользователей и экспертов с использованием любой модели машинного обучения.

Тем не менее, важно отметить, что основной проблемой, с которой мы столкнулись при повышении объяснимости табличных данных, было отсутствие базы знаний и экспертных знаний. В сравнении с наборами изображений, где «объяснительные» свойства предоставлены экспертами, для табличных наборов данных может потребоваться тщательный анализ с целью обнаружения новых значимых закономерностей в массиве информации. Наш эксперимент проводился на небольшом наборе данных, однако мы полагаем, что разработка надлежащей онтологии является сложной задачей в условиях больших и комплексных наборов данных. Полученные результаты свидетельствуют о том, что поиск и построение новых свойств может потребовать значительных ресурсов и глубокого анализа. Построение корректных свойств непосредственно влияет на процесс логического вывода, поскольку формирование исчерпывающих правил SWRL может происходить только при наличии глубоко детализированных «объяснительных» свойств.

Также хотелось бы отметить оценку качества разработанной нами XAI системы. В отличие от других XAI систем, где основной метрикой может являться точность предсказаний, именно для нашего подхода и нашей объяснимой системе критически важны три аспекта качества: интерпретируемость, надежность и способность обрабатывать сложную семантическую информацию. Для оценки интерпретируемости нашей системы мы использовали качественный анализ полноты и понятности генерируемых объяснений. Разработанные объяснения оценивались по следующим

критериям: (1) полнота раскрытия причинно-следственных связей, (2) использование релевантных «объяснительных» свойств, и (3) адаптация к ментальным моделям пользователей различного уровня компетенции. Надежность системы оценивалась через согласованность объяснений при небольших вариациях во входных данных. Благодаря использованию нечеткой логики система демонстрирует стабильность объяснений даже при наличии шума или неопределенности в данных, что является существенным преимуществом перед бинарными системами классификации. Способность системы обрабатывать сложную семантическую информацию была протестирована через интеграцию онтологических знаний с нечеткими данными. Наш подход позволяет эффективно связывать количественные показатели (степени принадлежности) с качественными экспертными знаниями, структурированными в онтологии, что создает более полную картину для пользователя и снижает когнитивную нагрузку при интерпретации результатов.

Заключение

В этом исследовании анализируется новый предложенный подход, который дает представление о том, как концепция «объяснительных» свойств может помочь улучшить XAI за счет повышения его объяснимости, используя нечеткую кластеризацию. Новая методика полностью соответствует архитектуре, предложенной Bellucci и др. для классификатора изображений на основе онтологий (OBIC). Также используется нечеткая кластеризация для классификации данных. В результате получена XAI система, которая повышает интерпретируемость и достоверность объяснений, посредством нечеткого подхода, интегрированного с «объяснимыми» свойствами в онтологии. Это исследование намеренно не оценивает и не оптимизирует точность базовых моделей или прогнозов так как основное внимание уделяется исключительно объяснительным аспектам.

Чтобы оценить наш подход, мы протестировали его на наборе данных для клинических записей о сердечной недостаточности. В результате эксперимента мы создали новую методику, которая позволяет оптимально использовать эти свойства и нечеткий подход, повышая достоверность объяснений прогнозирования и объяснимость табличных наборов данных. Выявленные свойства позволили нам гибко настроить онтологию и экспертные знания для соответствия концепции “объяснительных» свойств за счет улучшения объяснимости, а нечеткая логика решила проблему неточности данных, полученных из условий реального мира.

Мы планируем улучшить глубину и гибкость объяснений систем XAI, используя в будущем расширение нечеткой логики Z-числа, которые так же, как и нечеткая логика предложены Л.А. Заде. По словам Л.А. Заде, так называемый метод «вычисления словами» (англ. computing with words) обладает большим потенциалом в условиях неточных данных. Эта парадигма может обеспечить связь между вычислениями и естественным языком. Кроме того, в его другой работе была представлена возможность применения нечетких знаний для описания идеи семантики естественного языка. Заде заявил, что Z-числа полностью соответствуют этим двум идеям. Исходя из предложений Л.А. Заде, мы планируем провести дальнейшие эксперименты с «объяснительными» свойствами, объединив их в работе с Z-числами.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Dwivedi R., Dave D., Naik H., et al. Explainable AI (XAI): Core Ideas, Techniques, and Solutions. *ACM Computing Surveys*. 2023;55(9):1–33. https://doi.org/10.1145/3561048

2. Jo T. *Machine Learning Foundations: Supervised, Unsupervised, and Advanced Learning*. Cham: Springer; 2021. 391 p. https://doi.org/10.1007/978-3-030-65900-4
3. Saranya A., Subhashini R. A Systematic Review of Explainable Artificial Intelligence Models and Applications: Recent Developments and Future Trends. *Decision Analytics Journal*. 2023;7. https://doi.org/10.1016/j.dajour.2023.100230
4. Gunning D., Aha D.W. DARPA's Explainable Artificial Intelligence (XAI) Program. *AI Magazine*. 2019;40(2):44–58. https://doi.org/10.1609/aimag.v40i2.2850
5. Bellucci M., Delestre N., Malandain N., Zanni-Merk C. Combining an Explainable Model Based on Ontologies with an Explanation Interface to Classify Images. *Procedia Computer Science*. 2022;207:2395–2403. https://doi.org/10.1016/j.procs.2022.09.298
6. Kulmanov M., Smaili F.Z., Gao X., Hoehndorf R. Semantic Similarity and Machine Learning with Ontologies. *Briefings in Bioinformatics*. 2021;22(4). https://doi.org/10.1093/bib/bbaa199
7. Giustozzi F., Saunier J., Zanni-Merk C. A Semantic Framework for Condition Monitoring in Industry 4.0 based on Evolving Knowledge Bases. *Semantic Web*. 2023;15(3):1–29. https://doi.org/10.3233/SW-233481
8. Bourgais M., Giustozzi F., Vercouter L. Detecting Situations with Stream Reasoning on Health Data Obtained with IoT. *Procedia Computer Science*. 2021;192:507–516. https://doi.org/10.1016/j.procs.2021.08.052
9. Zadeh L.A. Fuzzy Sets. *Information and Control.* 1965;8(3):338–353. https://doi.org/10.1016/S0019-9958(65)90241-X
10. Aliev R.A., Aliev R.R. *Soft Computing and Its Applications*. Singapore: World Scientific; 2001. 460 p. https://doi.org/10.1142/4766
11. Dumitrescu C., Ciotirnae P., Vizitiu C. Fuzzy Logic for Intelligent Control System Using Soft Computing Applications. *Sensors*. 2021;21(8). https://doi.org/10.3390/s21082617
12. Gardashova L.A. Synthesis of Fuzzy Terminal Controller for Chemical Reactor of Alcohol Production. In: *10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions – ICSCCW-2019, 27–28 August 2019, Prague, Czech Republic*. Cham: Springer; 2020. P. 106–112. https://doi.org/10.1007/978-3-030-35249-3_13
13. Kosov P., El Kadhi N., Zanni-Merk C., Gardashova L. Advancing XAI: New Properties to Broaden Semantic-Based Explanations of Black-Box Learning Models. *Procedia Computer Science*. 2024;246:2292–2301. https://doi.org/10.1016/j.procs.2024.09.560
14. Bezdek J.C., Ehrlich R., Full W. FCM: The Fuzzy *C*-Means Clustering Algorithm. *Computers & Geosciences*. 1984;10(2–3):191–203. https://doi.org/10.1016/0098-3004(84)90020-7
15. Kosov P., El Kadhi N., Zanni-Merk C., Gardashova L. Semantic-Based XAI: Leveraging Ontology Properties to Enhance Explainability. In: *2024 International Conference on Decision Aid Sciences and Applications (DASA), 11–12 December 2024, Manama, Bahrain*. IEEE; 2025. P. 1–5. https://doi.org/10.1109/DASA63652.2024.10836289
16. Jones N.A., Ross H., Lynam T., Perez P., Leitch A. Mental Models: An Interdisciplinary Synthesis of Theory and Methods. *Ecology and Society*. 2011;16(1). URL: http://www.jstor.org/stable/26268859
17. Horrocks I., Patel-Schneider P.F., Boley H., Tabet S., Grosof B., Dean M. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. World Wide Web Consortium. URL: https://www.w3.org/submissions/SWRL [Accessed 12th March 2025].

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Косов Павел Игоревич, аспирант кафедры компьютерной инженерии, Азербайджанский государственный университет нефти и промышленности, Баку, Азербайджан.
e-mail: pavel_kosov@asoiu.edu.az
ORCID: 0009-0005-5602-2086

Pavel I. Kosov, Postgraduate at the Department of Computer Engineering, Azerbaijan State Oil and Industry University, Baku, Azerbaijan.

Гардашова Латафат Аббас гызы, доктор технических наук, профессор, проректор по научной работе, Азербайджанский государственный университет нефти и промышленности, Баку, Азербайджан.
e-mail: l.qardashova@asoiu.edu.az
ORCID: 0000-0003-3227-2521

Latafat A. Gardashova, Doctor of Engineering Sciences, Professor, Vice-Rector for Scientific Affairs, Azerbaijan State Oil and Industry University, Baku, Azerbaijan.

Статья поступила в редакцию 27.03.2025; одобрена после рецензирования 18.04.2025; принята к публикации 24.04.2025.

The article was submitted 27.03.2025; approved after reviewing 18.04.2025; accepted for publication 24.04.2025.