ИНТЕРПРЕТИРУЕМАЯ РЕКОНСТРУКЦИЯ НЕЙТРОННЫХ СПЕКТРОВ НА ОСНОВЕ ДВУХЭТАПНОГО ОБУЧЕНИЯ ANFIS С SHAP-РЕГУЛЯРИЗАЦИЕЙ

Чижов К.А., Лебедев А.Д.[1], Трофимов Ю.В.[2], Ильин А.С.[1], Лебедев М.Д.[3]

Объединённый институт ядерных исследований, Россия, 141980, г. Дубна, ул. Жолио-Кюри, 6; kchizhov@jinr.ru

[1]Государственный университет «Дубна», Россия, 141980, г. Дубна; lad.24@uni-dubna.ru, a.ilin@innopolis.university

[2]ЛИТ им. М.Г. Мещерякова ОИЯИ, Россия, 141980, г. Дубна, ул. Жолио-Кюри, 6; ura_trofim@bk.ru

[3]НИТУ «МИСиС», Россия, 119049, г. Москва; lebedevmisha2003@yandex.ru

rus]mce

Восстановление энергетического спектра нейтронного излучения по показаниям детекторов Боннера представляет собой некорректно поставленную обратную задачу и критично для радиационной безопасности, дозиметрии и экспериментальной физики. Классические методы развёртки чувствительны к априорным предположениям, а нейросетевые подходы часто выступают в роли «чёрного ящика». Нечётко-нейронные модели типа ANFIS обладают прозрачной структурой правил, но без специальной настройки могут использовать физически неинтерпретируемые зависимости и застревать в локальных минимумах.

В работе рассматривается задача реконструкции спектра по десяти каналам сфер Боннера (вектор входов $\mathbf{x} \in \mathbb{R}^{10}$) с предсказанием распределения по 60 энергетическим бинам ($\mathbf{y} \in \mathbb{R}^{60}$). В качестве базовой модели используется адаптивная нейро-нечёткая система вывода ANFIS с гауссовскими функциями принадлежности и линейной частью в заключительном слое. Предлагается двухэтапный конвейер обучения, сочетающий глобальную оптимизацию параметров и структуры с внедрением объяснимых ограничений на использование признаков.

На первом этапе (Vanilla) параметры ANFIS подбираются методом роя частиц (PSO). Этот этап обеспечивает устойчивую инициализацию в многомодальном пространстве параметров и формирует осмысленную базу нечетких правил. На втором этапе выполняется градиентная донастройка (fine-tune) с использованием SHAP-подобных важностей признаков в качестве регуляризатора: к основной MSE-функции добавляется штраф на профиль SHAP-весов.

Экспериментальные исследования проведены на реальных данных нейтронной спектрометрии с использованием сфер Боннера. В качестве базового ориентира рассматривается «ванильная» модель (первый этап без SHAP-регуляризации), достигающая $R^2_{\text{weighted}} \approx 0{,}86$ на тестовой выборке. Предлагаемый двухэтапный метод с SHAP-регуляризацией демонстрирует сопоставимую точность ($R^2_{\text{weighted}} \approx 0{,}84$) при наличии дополнительных свойств интерпретируемости. Таким образом, интерпретируемость становится внутренним свойством обученной модели, а не только постфактум-анализом «чёрного ящика».

Работа выполнена в рамках государственного задания Министерства науки и высшего образования Российской Федерации (тема № 124112200072-2).