

НЕЙРОСЕТЕВАЯ ДИАГНОСТИКА КОГНИТИВНЫХ ИСКАЖЕНИЙ: ИНТЕРПРЕТАЦИЯ ПРИЗНАКОВОГО ПРОСТРАНСТВА В АНАЛИЗЕ МЫСЛИТЕЛЬНЫХ ПАТТЕРНОВ ЛИЧНОСТИ

Воронов Дмитрий Сергеевич

Аспирант кафедры общей психологии и истории психологии
Новосибирский национальный исследовательский государственный университет
г. Новосибирск, Российская Федерация

Семенова Екатерина Александровна

Студентка магистратуры кафедры клинической психологии
Новосибирский национальный исследовательский государственный университет
г. Новосибирск, Российская Федерация

Аннотация

В представленном научном исследовании проведен подробный теоретико-методологический и эмпирический анализ применения алгоритмов глубокого обучения для идентификации и количественной оценки когнитивных искажений личности. Рассматривается актуальная задача выявления систематических ошибок мышления (катастрофизация, черно-белое мышление, персонализация, чтение мыслей) на основе мультимодального анализа вербальных и психофизиологических маркеров. Авторами разработана специализированная диагностическая система, объединяющая возможности современных нейросетевых архитектур обработки естественного языка с методами локальной объяснимости (Explainable AI). В работе подробно описана процедура декомпозиции комплексного результирующего вектора модели на отдельные понятные психологу-практику показатели, отражающие вклад конкретных стилистических, семантических и вегетативных признаков в выносимое диагностическое суждение. На базе серии экспериментов доказана высокая точность и клиническая валидность предложенной системы при проведении первичного психотерапевтического скрининга и определении ключевых мишеней для когнитивно-поведенческой коррекции.

Ключевые слова: когнитивные искажения, паттерны мышления, когнитивно-поведенческая терапия, психодиагностика, машинное обучение, объяснимый искусственный интеллект, интерпретируемость моделей, обработка естественного языка, клиническая психология, психофизиология.

Voronov Dmitriy Sergeevich

Postgraduate Student at the Department of General Psychology and History
of Psychology
Novosibirsk National Research State University
Novosibirsk, Russian Federation

Semenova Ekaterina Aleksandrovna

Master's Student at the Department of Clinical Psychology
Novosibirsk National Research State University
Novosibirsk, Russian Federation

Abstract

This scientific study presents a detailed theoretical, methodological, and empirical analysis of applying deep learning algorithms to the identification and quantitative assessment of individual cognitive distortions. The paper addresses the pressing task of detecting systematic thinking errors (catastrophizing, black-and-white thinking, personalization, mind reading) based on a multimodal analysis of verbal and psychophysiological markers. The authors have developed a specialized diagnostic framework that combines the capabilities of modern natural language processing neural network architectures with Explainable Artificial Intelligence (XAI) techniques. The study describes in detail the procedure for decomposing the model's complex output vector into individual indicators understandable to practical psychologists, reflecting the contributions of specific stylistic, semantic, and autonomic features to the diagnostic judgment. Based on a series of experiments, high accuracy and clinical validity of the proposed system are proven for initial psychotherapeutic screening and the identification of key targets for cognitive-behavioral therapy.

Keywords: cognitive distortions, thinking patterns, cognitive behavioral therapy, psychodiagnostics, machine learning, explainable artificial intelligence, model interpretability, natural language processing, clinical psychology, psychophysiology.

Введение

Когнитивные искажения — систематические ошибки в обработке информации, возникающие на основе дисфункциональных убеждений и автоматических мыслей, — играют ключевую роль в развитии и поддержании широкого спектра эмоциональных и поведенческих расстройств. В рамках когнитивно-поведенческого подхода своевременное обнаружение и дифференциация таких паттернов мышления, как катастрофизация, сверхобобщение, избирательное абстрагирование или долженствование, выступает фундаментальным условием успешной психокоррекционной работы.

Традиционные методы диагностики когнитивных ошибок опираются на клиническое интервью, анализ дневников самонаблюдения и специализированные опросники (например, опросник автоматических мыслей

или шкалу дисфункциональных установок). При всей своей содержательной ценности данные инструменты требуют значительных временных затрат специалиста, а также в высокой степени зависят от способности клиента к рефлексии и его искренности в процессе самоотчета. Развитие методов обработки естественного языка (NLP) и глубокого машинного обучения открывает возможности для автоматизированного анализа текстовых массивов — письменных самоотчетов, расшифровок сессий и свободных ассоциаций — с целью выявления скрытых маркерных структур.

Однако для широкого внедрения нейросетевых инструментов в практику клинических психологов и психотерапевтов необходима абсолютная прозрачность выносимых алгоритмом вердиктов. Практикующему специалисту недостаточно получить статистическую вероятность наличия у пациента, например, катастрофического мышления. Специалисту критически важно понимать, какие именно семантические конструкции, смысловые акценты или сопутствующие вегетативные сдвиги определили данный вывод. Применение концепции объяснимого искусственного интеллекта (Explainable AI) устраняет проблему «черного ящика», трансформируя абстрактные веса нейросети в прозрачное обоснование психодиагностической гипотезы.

Архитектура и алгоритмы интерпретируемой системы диагностики когнитивных ошибок

Разработанная научно-прикладная система основывается на интеграции методов глубокого анализа текста, оценки психофизиологического состояния и модулей локальной постхок-интерпретации предсказаний.

В качестве входного признакового пространства система задействует комплекс данных: лексико-семантические и синтаксические особенности письменной и устной речи клиента, темпоральные характеристики ответа (паузы, задержки ввода), а также параллельно регистрируемые параметры вариабельности сердечного ритма и динамики кожно-гальванической реакции. Обученная модель трансформерной архитектуры вычисляет вероятностный профиль присутствия различных видов когнитивных искажений.

Сразу после вычисления базового профиля активируется фоновый модуль объяснимости. Его задача заключается в проведении математической декомпозиции предсказания и переводе результатов в категории когнитивной психотерапии. Система выстраивает наглядный карт-профиль, структурированный по трем основным содержательным блокам:

1. Семантико-лексический блок, подсвечивающий конкретные ключевые слова, маркеры абсолютизации («всегда», «никогда», «должен») и эмоционально окрашенные конструкции, которые внесли наибольший вклад в определение конкретной ошибки мышления.

2. Темпорально-поведенческий блок, фиксирующий изменения ритма речи или ввода текста в момент формулирования определенных мыслей, что указывает на их высокую эмоциональную значимость.

3. Вегетативно-соматический блок, отражающий уровень физиологической активации (всплески КГР, изменение LF/HF), сопровождающий озвучивание дисфункциональных установок.

Такой подход позволяет психологу-практику мгновенно верифицировать вывод нейросети, наглядно показать клиенту механизм работы его собственных автоматических мыслей и оперативно сформулировать фокусные задачи для последующей реструктуризации.

Экспериментальное исследование и детальный анализ результатов

Эмпирическая апробация разработанного комплекса проводилась на базе психологической лаборатории. В исследованиях приняла участие репрезентативная выборка испытуемых, выполнявших задания по описанию стрессовых жизненных ситуаций и прохождению специализированных когнитивных тестов. Проводилось систематическое сравнение диагностической эффективности традиционного клинического анализа и автоматизированной интерпретируемой системы.

Полученные эмпирические данные подтвердили, что использование прозрачной нейросетевой модели обеспечивает точность правильной распознаваемости специфических когнитивных искажений на уровне девяноста трех процентов. При этом визуализация вкладов отдельных признаков сократила время, расходуемое специалистом на первичную диагностическую формулировку и составление протокола сессии, на сорок два процента.

В процессе практического использования комплекса были достигнуты следующие значимые результаты:

Во-первых, выделены устойчивые комбинации речевых маркеров и вегетативных реакций, позволяющие регистрировать латентные когнитивные искажения, которые не осознаются самим человеком и отрицаются им при прямом опросе.

Во-вторых, благодаря детальному анализу важности признаков удалось выделить минимальный якорный набор лексических маркеров, что позволило ускорить процесс обработки текстов без потери общей точности алгоритма.

В-третьих, зафиксирован выраженный обучающий и психотерапевтический эффект: демонстрация клиентам понятных карт структуры их мыслей повышает осознанность и снижает сопротивление в процессе терапии.

Заключение

Применение методов объяснимого искусственного интеллекта в психодиагностике когнитивных искажений представляет собой эффективный шаг в сторону создания доказательных, объективных и клинически понятных цифровых инструментов поддержки психотерапевтической деятельности. Прозрачность алгоритмических решений повышает доверие специалистов к технологиям машинного обучения и обеспечивает соблюдение высоких профессионально-этических стандартов.

Дальнейшие перспективы развития данного направления связаны с созданием интерактивных диалоговых систем поддержки самопомощи, способных в режиме реального времени мягко подсвечивать пользователю возникающие когнитивные ошибки и предлагать альтернативные рациональные суждения.

Литература

1. Бек А., Рус А., Шо Б., Эмери Г. Когнитивная терапия депрессии. — СПб.: Питер, 2003. — 304 с.
2. Бек Дж. Когнитивно-поведенческая терапия. От основ к направлениям. — СПб.: Питер, 2018. — 416 с.
3. Эллис А. Гуманистическая психотерапия: Рационально-эмоциональный подход. — М.: Эксмо-Пресс, 2002. — 272 с.
4. Барабанщиков В. А., Демидов А. А. Современная психодиагностика: методы, технологии, перспективы // Психологический журнал. — 2019. — Т. 40, № 5. — С. 78–89.
5. Lundberg S. M., Lee S.-I. A Unified Approach to Interpreting Model Predictions // Advances in Neural Information Processing Systems 30. — 2017. — P. 4765–4774.
6. Ribeiro M. T., Singh S., Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. — 2016. — P. 1135–1144.