

Оценивание креативности на основе анализа изображений с помощью нейронных сетей *

И. Л. Угланова, Е. С. Гельвер, С. В. Тарасов, Д. А. Грачева, Е. Е. Вырва

Национальный исследовательский университет «Высшая школа экономики», г. Москва, Россия

Аннотация. В статье исследуются возможности оценивания креативности обучающихся на основе нейросетевых подходов к анализу изображений. Применение психометрического анализа данных в методологии латентного классового анализа позволило получить разметку для обучения нейронной сети без привлечения экспертов. Высокая точность предсказаний сети для идентификации креативности изображений позволяет говорить о масштабных перспективах машинного обучения для оценки сложных образовательных и психологических характеристик.

Ключевые слова: креативность, анализ изображений, нейронные сети, образовательное оценивание, психометрика, машинное обучение.

DOI 10.14357/20718594210109

Введение

На сегодняшний день оценка сложных, метапредметных навыков является актуальной задачей системы образования. Примером такого навыка выступает креативность, на оценке которой и будет сфокусирована данная работа. Креативность включена в большинство рамок ключевых грамотностей и навыков современного мира: доклад Всемирного экономического форума [1], исследовательские работы [2, 3] и российские образовательные стандарты [4, 5]. Для оценивания сложных навыков требуется прибегать не только к традиционным форматам тестирования (задания с выбором ответа или эссе), но и к более современным форматам.

Часто инновационные форматы предлагают среду тестирования, в которой требуется графическое решение тестового задания. Это особенно актуально для заданий на креативность,

где тестируемых просят создать некоторый продукт. Оценивание продуктов деятельности тестируемых часто проводится с привлечением экспертов. При этом экспертное оценивание обладает рядом существенных недостатков (ресурсозатратность на обучение экспертов, субъективные искажения, отложенная обратная связь и др.) [6]. Кроме того, для валидной оценки заданий с графическим решением недостаточно учитывать каждое конкретное действие тестируемого, а требуется учет целостности продукта, созданного в ходе тестирования. Для повышения качества оценивания таких заданий все более востребованным становится использование методов оценки на основе машинного обучения.

Такие технологии, как машинное обучение или искусственный интеллект, активно развиваются во многих областях, в том числе и в образовании [7]. Например, автоматический анализ текстов в образовательном тестировании

* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта №19-29-14110.

✉ Тарасов Сергей Владимирович. E-mail: svtarasov@hse.ru

уже стал популярным инструментом в исследованиях и практике [8-11]. Автоматический анализ изображений по сравнению с анализом текстов исследован не так подробно и применяется не так широко. В то же время один из самых распространенных подходов к оцениванию креативности – это оценивание через анализ изображений (традиционно с привлечением экспертов) [12].

Данное исследование представляет компьютеризованную систему оценивания креативности с помощью автоматического анализа изображений, чтобы избежать субъективности, вносимую экспертами, а также получить другие преимущества компьютерного тестирования – незамедлительную обратную связь, экономию временных и денежных ресурсов за счет отказа от работы экспертов.

Цель данной работы – продемонстрировать возможности машинного обучения для анализа изображений без привлечения экспертов в образовательном тестировании. В этом исследовании мы фокусируемся на оценивании креативности.

1. Опыт применения машинного обучения при анализе изображений

Изображения занимают значительную часть в массиве больших данных. Например, матрицы ДНК, астрономические наблюдения, спутниковые карты, трехмерное магнитно-резонансное изображение или компьютерная томография. Для изучения таких источников данных требуются сложные инструменты, которые должны быть удобными в использовании. Так, в одном из исследований описывается набор алгоритмов обработки изображений, реализованных на языке программирования Python с открытым исходным кодом [13].

В образовательном тестировании необходимость применения машинного обучения для анализа изображений вызвана тем, что тест может подразумевать графическое решение задания. В недавнем исследовании применялась технология обработки изображений, где изображения служили аргументами при решении научной задачи [14]. С помощью статистического анализа была выявлена взаимосвязь между характеристиками изображений и уровнем

успеваемости учащихся при написании научных текстов.

Графическое решение задач можно встретить в статьях по изучению навыков наложения швов хирургами [15, 16]. Авторы обнаружили, что простой компьютерный алгоритм может извлекать переменные из цифровых изображений шва и быстро обеспечивать количественную итоговую оценку обратной связи. Этот подход используется как инструмент обучения с объективной оценкой качества наложения швов – основного навыка для многих медицинских специальностей. В другой научной работе показано, что метрики, полученные в результате анализа изображений, оказались статистически значимыми в дифференциации навыков шитья между опытными и начинающими хирургами [16].

Область применения автоматической обработки изображений явно может быть расширена для использования в других инструментах психологического и образовательного тестирования. Например, для измерения таких навыков, как воображение [17], и креативность [18, 19]. Обратим внимание, что в российских исследованиях психологии и образования область автоматического анализа изображений только начинает развиваться [20].

2. Методология

2.1. Инструмент, процедура сбора данных, выборка

В данной работе мы рассматриваем инструмент для оценки навыков 21-го века 4К среди учеников 4 класса, разработанный сотрудниками Центра психометрики и измерений в образовании (Лаборатория измерения новых конструкций и дизайна тестов) Института образования НИУ ВШЭ¹.

В данный инструмент входит шесть заданий сценарного типа, оценивающих четыре навыка: критическое мышление, креативность, коммуникацию и кооперацию. В целом, в инструменте 4К существует тенденция, что одно задание позволяет измерить несколько навыков. В рамках данной работы рассматривалось одно из

¹ в рамках договора о научно-исследовательской работе по разработке инструмента оценки навыков «4К» с фондом «Вклад в будущее»

шести заданий под названием «Монстр», которое в большей степени направлено на оценку навыка креативности.

Инструмент предьявляется тестируемому в компьютерной форме. Экран заданий интерактивен: тестируемый нажимает на выбранную им область, после чего видит заранее подготовленную специфическую реакцию системы тестирования. Такой формат позволяет проявить сложные навыки, а также поддерживает мотивацию тестируемых и снижает тестовую тревожность. В то же время имитация реальной среды позволяет более точно зафиксировать наблюдаемое поведение, то есть свидетельство того, что тестируемый обладает конкретным навыком.

Для создания инструмента измерения разработчиками была определена теоретическая рамка конструкта креативности [21], которая основывается на классических и современных теориях [3, 12, 22, 23]. Креативность в данном исследовании понимается как способность представить и разработать принципиально новые подходы к решению проблем, ответы на вопросы, стоящие перед субъектом, или способы выражения идей для решения задач в повседневной жизни. Креативность включает два субконструкта:

- *Оригинальность* – способность продуцировать новые идеи и решения задач, которая может выражаться в появлении новых идей, в установлении новых, ранее не фиксировавшихся, связей между существующими идеями.
- *Детальность* – способность глубоко, с большой степенью подробности, проработать предложенную идею.

В задании «Монстр» каждый тестируемый, используя элементы конструктора, создает трех монстров, которые, по его мнению, «удивительны, необычны и отличаются от остальных монстров в городе». Элементы конструктора разложены на несколько групп (туловище, рот, глаза и т.д.), каждый элемент можно использовать несколько раз. Интерфейс задания представлен на Рис. 1. Несколько монстров были представлены на рабочем экране и считались референсами, относительно которых рассматривалась оригинальность построенных тестируемым монстров.

Для каждого тестируемого сохраняются значения индикаторов по трем построенным монстрам. Для оценки оригинальности использовались шесть индикаторов, оцениваемых дихотомически, где значение «1» присваивалось, если:

- расположение ног отлично от референса (не в нижней части тела);
- расположение рук отлично от референса (не в верхней части тела);
- расположение рта отлично от референса (не в нижней части туловища);
- расположение глаз отлично от референса (не в верхней части туловища);
- использовано больше 1 туловища;
- не наблюдается симметрия конечностей (рук и ног).

Для оценки детальности использовались два дихотомических индикатора: значение «1» присваивалось, если был изменен цвет монстра и если было применено вращение к элементам монстра; и семь индикаторов на общее количество элементов и количество элементов из каждой группы.

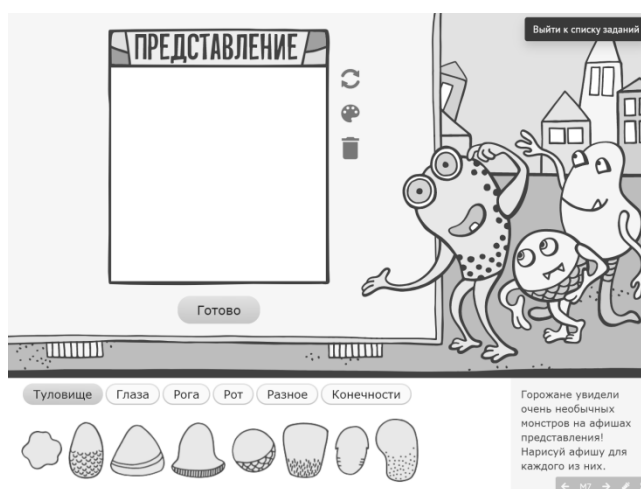


Рис. 1. Интерфейс рабочей области построения монстра в задании «Монстр»

В исследовании принимали участие 1780 учащихся четвертых классов из трех городов России – Москвы, Калуги и Ярославля. Среднее время выполнения задания «Монстр» – 10 минут. Так как каждый тестируемый строил по три монстра, то исходная база содержала 5340 изображений. После удаления профилей с пропущенными значениями по всем индикаторам количество монстров в базе сократилось до 5286 (1762 профиля учащихся).

2.2. Методология анализа данных

Анализ данных проводится в два последовательных этапа. Первый этап служит цели создать разметку по классам для обучения нейросети. Источником информации для разметки выступают результаты психометрического анализа данных в методологии латентного классового анализа (Latent Class Analysis, LCA). В данном исследовании мы оцениваем тестируемых по продукту их деятельности (созданным монстрам) по двум субконструктам креативности (оригинальность и детальность) отдельно. На втором этапе происходит выбор архитектуры и обучение нейросети.

1 этап. Латентный классовый анализ является частью методологии смешанных моделей (Mixture Modeling) и позволяет по группе индикаторов выявить латентные классы тестируемых. В отличие от кластерного анализа, в основе LCA лежит математическая модель, которая позволяет оценить, насколько хорошо предлагаемая модель классификации подходит данным.

Пусть есть M наблюдаемых бинарных индикаторов. Модель предполагает, что все M наблюдаемых индикаторов образуют ненаблюдаемую категориальную переменную c , состоящую из K классов.

Модель LCA включает два ключевых параметра – размер латентных классов и условную вероятность индикаторов. Размер латентного класса p_k определяет пропорцию тестируемых в классе k , при этом каждый тестируемый приписан только к одному классу, то есть $\sum_{i=1}^k p_k = 1$. Условные вероятности индикаторов t_{mk} описывают связь между значением наблюдаемого индикатора m и оценкой категориальной латентной переменной c (по аналогии с факторными нагрузками в методологии факторного анализа). Другими словами, условная вероятность индикатора определяет вероят-

ность тестируемого из латентного класса k быть отнесенным к определенной категории наблюдаемого индикатора (в бинарном случае две категории, где категория 1 – наличие признака оцениваемого конструкта, категория 0 – отсутствие признака оцениваемого конструкта).

Сравнение классов проводится как по значениям условных вероятностей, так и по показателям отношения шансов принадлежности к классу (odds-ratio). Значение отношения шансов показывает, во сколько раз шансы обнаружить признаки оцениваемого конструкта выше в одном классе по сравнению с другим (контрольным). Пусть существует два класса, где первый класс – контрольный. Тогда если отношения шансов превышают 1, то шансы обнаружить признаки оцениваемого конструкта больше во втором классе по сравнению с первым. В случае, когда отношение шансов меньше 1, шансы обнаружить признаки оцениваемого конструкта больше в первом классе.

Для того, чтобы определить структуру модели LCA, то есть выявить оптимальное количество латентных классов с различным уровнем выраженности конструкта, проводится разведывательный (эксплораторный) анализ. Строится серия моделей с разным количеством классов (количество классов определяется исследователем).

Выбор лучшей модели осуществляется по следующим критериям. Качество классификации наблюдений определяется по показателю энтропии, где значения больше 0,8 считаются приемлемыми [24]. Сравнение моделей с разным количеством классов происходит по информационному критерию Акаике (AIC) и байесовскому информационному критерию (BIC) в пользу моделей, где значения критериев минимальны. Дополнительно, тест Лу-Менделя-Рубина (Lo-Mendell-Rubin Adjusted Likelihood Ratio Test, LRT) используется для сравнения моделей с количеством классов, отличающихся на один. Значимость этого критерия свидетельствует о том, что модель с меньшим количеством классов подходит данным хуже. Согласно рекомендации [25], отдается предпочтение моделям, где минимальная пропорция наблюдений в одном классе составляла не менее 8%.

В данной работе количество латентных классов характеризует количество классов монстров с различным уровнем выраженности ори-

гинальности и детальности. Результаты LCA будут служить начальной разметкой для обучения нейронной сети и применения машинного обучения. Анализ проведен в программе Mplus, версия 8.3 [26].

2 этап. Машинное обучение (Machine Learning) – это подраздел искусственного интеллекта, который находится на стыке математической статистики, методов оптимизации и классических математических дисциплин. Оно изучает построение алгоритмов, которые способны обучаться. Одной из самых распространенных задач в области машинного обучения является задача классификации. Этот класс задач относится к «обучению с учителем» (Supervised Learning), в котором по конечному множеству прецедентов необходимо выявить общую закономерность. Процесс идентификации может быть автоматизирован при помощи бинарного классификатора изображений.

Среди всех алгоритмов машинного обучения самыми эффективными являются методы глубокого обучения (Deep Learning). Глубокое обучение – совокупность методов машинного обучения, основанных на обучении представлениям, а не специализированным алгоритмам под конкретные задачи. «Глубина» обучения обусловлена многослойностью искусственных нейронных сетей.

В 1998 году Ян Лекун опубликовал работу, посвященную новой нейросетевой архитектуре [27], которая осталась незамеченной вплоть до 2012 года, когда были проведены соревнования по компьютерному зрению (Computer Vision) на основе данных ImageNet [28]. На этом соревновании была представлена сверточная нейронная сеть (Convolutional Neural Network, CNN), которая способна классифицировать миллионы изображений с ошибкой 15,2 %. Начиная с 2012 года, качество работы сверточных нейронных сетей постоянно возрастало. В результате оно стало сопоставимо с человеческими возможностями [29] (Рис. 2).

В основе работы сверточных сетей лежат фильтры, которые занимаются распознаванием определенных паттернов изображения. Фильтр состоит из ядер, являющихся матрицами чисел. Ядра перемещаются вдоль изображения и идентифицируют наличие искомого паттерна на конкретном участке. Идентификация происходит путем совершения операции свертки –

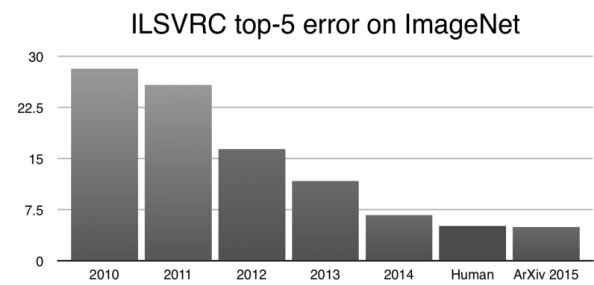


Рис. 2. Статистика ошибок, совершенных в процессе анализа данных ImageNet за 2010–2015 г.г.

суммированием произведений элементов фильтра и матрицы входных сигналов [30]. Большое значение свертки свидетельствует о наличии паттерна в искомой области. Между сверточными слоями располагаются субдискретизирующие слои (pooling layers, max pooling layers), которые отвечают за сжатие изображения при помощи сложения блоков пикселей.

В результате, после прохождения сверточных и субдискретизирующих слоев, изображение представляется как набор сверточных карт (карт признаков), который подключается к полносвязному слою. Обучение нейронной сети происходит путем обратного распространения ошибки (Back propagation), целью которого является регулирование веса каждого нейрона пропорционально его вкладу в общую ошибку.

Так как обучение нейронной сети с нуля является сложным вычислительным процессом, мы воспользуемся методом обучения Transfer Learning, который нашел широкое применение в задачах компьютерного зрения на медицинских снимках [31]. Transfer Learning – это метод обучения, когда «знания» (веса) нейросети, обученной на одной задаче, переносятся на другую задачу.

Сравнительные характеристики наиболее распространенных архитектур предобученных нейронных сетей представлены в Табл. 1.

Классификация такого рода изображений является нестандартной задачей, поэтому выбор модели был основан на количестве слоев. Среди всех вышеперечисленных моделей наибольшее количество слоев у модели Densenet201 [32]. Дополнительным ее преимуществом является относительно небольшой размер.

Изначально данная сеть обучалась для задачи, где количество классов было 1000. Нашей задачей является бинарная классификация,

Табл. 1. Сравнение архитектур предобученных нейросетей

Модель	Структура модели	Количество слоев	Размер, МБ
AlexNet	5 conv + 3 fc	8	238
VGG-19	16 conv + 3 fc	19	560
ResNet-152	151 conv + 1 fc	152	255
Densenet201	200 conv + 1 fc	201	80

*conv – сверточный слой, fc – полносвязный слой

поэтому количество нейронов в последнем слое должно быть 2. Также, могут быть добавлены ещё несколько полносвязных слоев перед последним слоем, состоящем из двух нейронов.

Перед применением методов машинного обучения для анализа изображений необходимо предварительно обработать данные – сбалансировать классы, чтобы модель лучше обучалась и была более стабильной. Для балансировки классов мы воспользовались методом аугментации данных (Data Augmentation), который позволяет создать дополнительные данные для обучения из уже имеющихся. Для изображений это реализуется при помощи поворота изображений, смены цвета или отражения по горизонтали.

3. Результаты

Результаты сравнения моделей LCA с разным количеством классов для субконструкта оригинальность представлены в Табл. 2.

Предпочтение было отдано двух-классовому решению с наивысшим показателем энтропии

модели (0,833). Как видно по таблице, при увеличении количества классов показатель энтропии снижался. Минимальные значения информационных критериев достигнуты в модели с четырьмя классами. Однако в трех-классовом и четырех-классовом решении образовались малонаполненные классы. Малонаполненный класс, во-первых, трудно интерпретировать содержательно, и, во-вторых, его использование для обучения нейронной сети малоэффективно.

Рассмотрим подробнее, изображения каких монстров попали в первый и второй класс. Показатели отношения шансов по каждому индикатору представлены в Табл. 3. Указание категория >1 свидетельствует о бинарном характере индикатора, где категория 1 свидетельствует о наличии признака креативности, категория 0 – об отсутствии. Значения отношения шансов больше 1 позволяют говорить о том, что у монстров во втором классе по сравнению с первым чаще выражены черты оригинальности по соответствующему индикатору. По результатам разбиения на два класса все монстры во втором

Табл. 2. Результаты сравнения качества моделей Оригинальности

Модель	Энтропия	Минимальная пропорция наблюдений в классе, %	AIC	BIC	Статистика LMR
C = 2	0,833	31	26075,438	26160,884	1509,007*
C = 3	0,825	6	25911,146	26042,602	175,369*
C = 4	0,573	5	25865,959	26043,425	58,217*

* – $p < 0.05$

Табл. 3. Показатели отношения шансов для модели Оригинальности

Индикатор	Отношение шансов
Расположение ног (категория >1)	1,56*
Расположение рук (категория >1)	3,43*
Расположение рта (категория >1)	26,27*
Расположение глаз (категория >2)	2,63*
Симметрия конечностей (категория >1)	1,58*

* – $p < 0,05$

Табл. 4. Результаты сравнения качества моделей Детальности

Модель	Энтропия	Минимальная пропорция наблюдений в классе, %	AIC	BIC	Статистика LMR
C=2	0,816	16	63230,339	63419,769	2530,415*
C=3	0,660	16	62504,974	62792,386	749,509*
C=4	0,605	15	62329,210	62714,603	204,056*

* – $p < 0.05$

классе имеют одно туловище (то есть значение индикатора, отражающего использование больше 1 туловища, для всех монстров группы равно нулю), поэтому показатель значения шансов для этого индикатора не приводится.

Монстры из первого класса (69% от всех монстров) отличаются низкой выраженностью оригинальности по всем индикаторам. Монстры из второго класса (31% от всех монстров) характеризуются отличием от референса, то есть нарушением симметрии конечностей, расположением глаз в нижней части монстра и пр. Таким образом, по всем индикаторам отношения шансов больше 1, что позволяет говорить о том, что монстры в первом классе «скорее не оригинальны», а монстры во втором классе «скорее оригинальны».

Аналогичный анализ был проведен для выявления латентных классов с различным уровнем детальности как субконструкта креативности. Все индикаторы, отражающие количество элементов из разных групп, были преобразованы в порядковую шкалу (от 0 до максимум 6) с учетом наполненности категорий.

Результаты сравнения моделей с разным количеством классов представлены в Табл. 4. Двух-классовое решение отвечает высокому качеству классификации по сравнению с другими моделями. В связи с этим мы остановили свой выбор на модели с двумя классами.

Чтобы дать классам характеристику, обратимся к показателю отношения шансов (Табл. 4). Отношения шансов приведены для каждого порога порядковых индикаторов. Порог – это точка, в которой вероятность выбора следующей категории становится равной вероятности выбора предыдущей. Например, показатель отношения шансов у индикатора Разное категория >2 (0.490) означает, что по сравнению с монстрами из первого класса монстры из второго класса более вероятно попадают во

вторую категорию (два элемента из группы разное использованы при создании монстра), чем в первую категорию (один элемент из группы разное использован при создании монстра). Аналогичная интерпретация сохраняется для остальных порогов.

Согласно Табл. 5, показатели отношения шансов для модели Детальности меньше 1. Следовательно, у монстров в первом классе по сравнению со вторым реже выражены черты детальности по соответствующему индикатору и порогу. Монстры из второго класса вероятнее содержат больше элементов, состоят из элементов разных групп и элементов, к которым применяли поворот, а также имеют цвет, отличный от референтного (серый).

Табл. 5. Показатели отношения шансов для модели Детальности

Индикатор	Отношение шансов
Цвет (категория >1)	0,253*
Поворот (категория >1)	0,088*
Разное (категория >1)	0,263*
Разное (категория >2)	0,490*
Конечности (категория >1)	0,038*
Конечности (категория >2)	0,054*
Конечности (категория >3)	0,081*
Конечности (категория >4)	0,067*
Конечности (категория >5)	0,152*
Глаза (категория >1)	0,011*
Глаза (категория >2)	0,127*
Рот (категория >1)	0,068*
Рога (категория >1)	0,19*
Рога (категория >2)	0,158*

* – $p < 0.05$

Таким образом, второй класс (84% от всех монстров) характеризовался высоким уровнем детальности по всем индикаторам (скорее детально) по сравнению с первым классом (16% от всех монстров), где уровень детальности у монстров ниже.

Полученные разбиения монстров на классы являются начальной разметкой для обучения нейронной сети и применения машинного обучения – второго этапа анализа.

Качество работы нейронной сети для идентификации оригинальности изображений было проверено на тестовом наборе данных, состоящим из 1243 изображений (622 – оригинальных, 621 – неоригинальных) (Табл. 6), при помощи матрицы ошибок (Confusion Matrix). Матрица ошибок – это таблица, которая позволяет визуализировать эффективность алгоритма классификации путем сравнения прогнозируемого значения целевой переменной с ее фактическим значением.

Из матрицы ошибок видно, что 151 изображение из 1243 были неправильно классифицированы. Так как классы достаточно сбалансированы, то точность (Accuracy) может быть использована, как метрика оценки классификатора:

$$\text{Accuracy} = (606 + 486) / (606 + 486 + 135 + 16) = 0,88$$

Также, была построена и обучена нейронная сеть для идентификации детальности изображений. Модель была апробирована на выборке, также состоящей из 1243 изображений (622 – детальных, 621 – не детальных) (Табл. 7). Точность на данной выборке составила 87,6%.

Табл. 6. Матрица ошибок на тестовом наборе данных (оригинальность)

		Предсказанные значения	
		1	0
Истинные значения	1	606	135
	0	16	486

Табл. 7. Матрица ошибок на тестовом наборе данных (детальность)

		Предсказанные значения	
		1	0
Истинные значения	1	603	135
	0	19	486

4. Обсуждение результатов и заключение

Целью нашего исследования было представить результаты анализа данных с применением машинного обучения для образовательного оценивания. В частности, представлены результаты автоматического анализа изображений как способа оценивания креативности учеников в конце начальной школы. Одной из проблем такого тестирования становится необходимость экспертной проверки, которая увеличивает ресурсозатратность проведения тестирования (время и деньги на обучение экспертов), не предполагает быстрой обратной связи по результатам тестирования и подвержена субъективным искажениям (усталость эксперта, эффект ореола и т.д.). Другой проблемой выступает «смешивание» оцениваемых характеристик тестируемого при экспертной оценке. Поскольку инструмент измерения креативности 4К оценивает креативность как состоящую из двух факторов (оригинальности и детальности), мы можем ожидать, что эксперты будут оценивать более детальные изображения как оригинальные и наоборот.

Автоматический анализ изображений с использованием машинного обучения позволяет преодолеть эти проблемы. Для успешного воспроизведения экспертной деятельности автоматизированными процессами машинного обучения необходимо обеспечить надлежащее качество данных и предоставить качественную начальную разметку для обучения сети.

Качество данных в образовательном тестировании прежде всего зависит от качества инструмента измерения. Предварительный анализ качества инструмента измерения доказал его хорошие психометрические характеристики [21].

Разметка для обучения сети в нашей работе была получена также без привлечения экспертов. Для этого была применена модель латентного классового анализа, в которой данными выступали наличие и косвенное указание на расположение объектов в интерфейсе при построении изображения. Для получения разметки для детальности анализировалось общее количество добавленных объектов на экран, поворот элементов и изменение цвета. Для оригинальности – признаки отличия от референса. Модель LCA продемонстрировала, что

оптимальным является разбиение на два класса по уровню выраженности детальности и оригинальности.

Разбиение на два класса по каждому фактору стало основой для обучения нейросети. Для данной задачи была выбрана предобученная нейросетевая модель Densenet [32]. Точность предсказания составила 0,88 для фактора Детальность и 0,88 для фактора Оригинальность. С учетом специфики характеристики, для которой проводилось оценивание, мы склонны считать этот результат достаточно высоким. В отличие от большинства работ, связанных с компьютерным зрением, где различия между классами являются очевидными и наличие того или иного паттерна говорит о принадлежности классу, в нашем исследовании видимой разницы между классами нет. В связи с этим выбор нейросетевой архитектуры осуществлялся итерационно. Помимо Densenet, были протестированы следующие нейросетевые модели: Inception, AlexNet, ResNet, VGG.

Среди ограничений работы отметим, во-первых, небольшой размер выборки для обучения нейронной сети, а, во-вторых, принадлежность трёх монстров к одному ученику. Таким образом, мы можем ожидать, что изображения находятся в большем согласии между собой, чем если бы каждое изображение принадлежало независимому создателю. Построение трёх изображений одним учеником – необходимое решение для получения точных оценок креативности, поэтому в дальнейших исследованиях, направленных не только на анализ отдельных изображений, но и на предоставление индивидуальной обратной связи тестируемым, мы считаем важным учитывать такой общий источник дисперсии. Отметим, что валидизация и обоснование теоретической рамки креативности не входило в фокус данного исследования.

Работа вносит вклад в развитие междисциплинарных исследований – наук о данных и наук об образовании. Использование методов машинного обучения в образовательном тестировании – совсем новая, но перспективная область для решения исследовательских и прикладных задач [33]. Как показывает наша работа, методы машинного обучения имеют масштабные перспективы для оценки сложных образовательных и психологических характе-

ристтик, которые на сегодняшний день являются частью образовательных стандартов [4, 5].

Литература

1. Forum W. E. New vision for education: Unlocking the potential of technology / W. E. Forum, British Columbia Teachers' Federation Vancouver, BC, 2015.
2. Griffin P., Care E. Assessment and teaching of 21st century skills: Methods and approach / P. Griffin, E. Care, Springer, 2014.
3. Lucas B., Claxton G., Spencer E. Progression in student creativity in school: First steps towards new forms of formative assessments 2013.
4. Приказ Министерства образования и науки РФ от 6 октября 2009 г. N 373 «Об утверждении и введении в действие федерального государственного образовательного стандарта начального общего образования».
5. Приказ Министерства образования и науки РФ от 17 декабря 2010 г. N 1897 «Об утверждении федерального государственного образовательного стандарта основного общего образования».
6. Chuang T.-Y., Liu E. Z.-F., Shiu W.-Y. Game-based creativity assessment system: the application of fuzzy theory // *Multimedia Tools and Applications*. 2015. № 21 (74). P. 9141–9155.
7. Sekeroglu B., Dimililer K., Tuncal K. Student performance prediction and classification using machine learning algorithms 2019. P. 7–11.
8. Shao, Z., Li, Y., Wang, X., Zhao, X., Guo, Y. Research on a new automatic generation algorithm of concept map based on text analysis and association rules mining // *Journal of Ambient Intelligence and Humanized Computing*. 2020. № 2 (11). P. 539–551.
9. Liu, O. L., Rios, J. A., Heilman, M., Gerard, L., Linn, M. C. Validation of automated scoring of science assessments // *Journal of Research in Science Teaching*. 2016. № 2 (53). P. 215–233.
10. Nehm R. H., Ha M., Mayfield E. Transforming biology assessment with machine learning: automated scoring of written evolutionary explanations // *Journal of Science Education and Technology*. 2012. № 1 (21). P. 183–196.
11. Leacock C., Chodorow M. C-rater: Automated scoring of short-answer questions // *Computers and the Humanities*. 2003. № 4 (37). P. 389–405.
12. Torrance E. P. Predictive validity of the torrance tests of creative thinking. // *The Journal of creative behavior*. 1972.
13. Van der Walt S. [et al.]. scikit-image: image processing in Python // *PeerJ*. 2014. (2). P. e453.
14. Pei B., Xing W., Lee H.-S. Using automatic image processing to analyze visual artifacts created by students in scientific argumentation // *British Journal of Educational Technology*. 2019. № 6 (50). P. 3391–3404.
15. Kil I. Development and Preliminary Validation of Image-enabled Process Metrics for Assessment of Open Surgery Suturing Skill 2019.
16. Frischknecht, A. C., Kasten, S. J., Hamstra, S. J., Perkins, N. C., Gillespie, R. B., Armstrong, T. J., & Minter, R. M. The objective assessment of experts' and novices'

- suturing skills using an image analysis program // *Academic Medicine*. 2013. № 2 (88). P. 260–264.
17. Ward T. B. Structured Imagination: the Role of Category Structure in Exemplar Generation // *Cognitive Psychology*. 1994. № 1 (27). P. 1–40.
 18. Jankowska D. M., Karwowski M. Measuring creative imagery abilities // *Frontiers in psychology*. 2015. (6). P. 1591.
 19. Kim S. H., Vincent L. C., Goncalo J. A. Outside advantage: Can social rejection fuel creative thought? // *Journal of Experimental Psychology: General*. 2013. № 3 (142). P. 605.
 20. Славутская Е.В., Абриков В.С., Славутский Л.А. Нейросетевой системный анализ уровней психологических характеристик // *Вестник ЧГУ*. 2016. №1. URL: <https://cyberleninka.ru/article/n/neyrosetevoy-sistemnyy-analiz-urovnevyyh-psihologicheskikh-harakteristik> (дата обращения: 03.12.2020).
 21. Углонова И. Л., Орел Е. А., Брун И. В. Измерение креативности и критического мышления в начальной школе // *Психологический журнал* 2020. Том 41. № 6 [Электронный ресурс]. Доступ для зарегистрированных пользователей. URL: <http://ras.jes.su/psy/s020595920011124-2-1> (дата обращения: 28.11.2020).
 22. Богоявленская Д.Б. Психология творческих способностей. Москва : Федоров, 2009. 342 с.
 23. Sternberg R. J. Handbook of creativity / R. J. Sternberg, Cambridge University Press, 1999.
 24. Clark S. L., Muthén B. Relating latent class analysis results to variables not included in the analysis / S. L. Clark, B. Muthén, 2009.
 25. Nylund-Gibson K., Choi A. Y. Ten frequently asked questions about latent class analysis. // *Translational Issues in Psychological Science*. 2018. № 4 (4). P. 440.
 26. Muthén L. K., Muthén B. O. Mplus user's guide / Los Angeles, CA: Muthén & Muthén, 2007.
 27. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. Gradient-based learning applied to document recognition // *Proceedings of the IEEE*. 1998. № 11 (86). P. 2278–2324.
 28. Krizhevsky A., Sutskever I., Hinton G. E. Imagenet classification with deep convolutional neural networks // *Communications of the ACM*. 2017. № 6 (60). P. 84–90.
 29. He, K., Zhang, X., Ren, S., & Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification 2015. P. 1026–1034.
 30. Deep Neural Networks Applications in Handwriting Recognition [Электронный ресурс] / URL: http://technodocbox.com/3D_Graphics/70716176-Deep-neural-networks-applications-in-handwriting-recognition.html (дата обращения: 16.11.20)
 31. Raghu, M., Zhang, C., Kleinberg, J., Bengio, S. Transfusion: Understanding transfer learning for medical imaging // *Advances in neural information processing systems*. 2019. P. 3347–3357.
 32. DENSENET [Электронный ресурс] / URL: https://pytorch.org/hub/pytorch_vision_densenet (дата обращения: 23.11.20)
 33. Von Davier A. A. [et al.]. Computational psychometrics approach to holistic learning and assessment systems // *Frontiers in Education*. 2019. №4. P. 69.

Углонова Ирина Львовна. Магистр. Область научных интересов: психометрика, измерения в психологии и образовании. E-mail: iuglanova@hse.ru.

Гельвер Евгений Сергеевич. Магистр, старший бизнес-аналитик в Glowbyte Consulting. Область научных интересов: машинное обучение, компьютерное зрение, обработка естественного языка. E-mail: evgelver@gmail.com.

Тарасов Сергей Владимирович. Магистр. Область научных интересов: измерения в психологии и образовании, прогресс в обучении. E-mail: svtarasov@hse.ru.

Грачева Дарья Александровна. Магистр. Область научных интересов: психометрика, сопоставимость результатов тестирований. E-mail: dgracheva@hse.ru.

Вырва Елена Евгеньевна. Магистр. Область научных интересов: тест-девелопмент, измерения в психологии и образовании. E-mail: evyrva@hse.ru.

Assessing Creativity Using Image Analysis with Neural Networks

I. L. Uglanova, E. S. Gelver, S. V. Tarasov, D. A. Gracheva, E. E. Vyryva

National Research University Higher School of Economics, Moscow, Russia

The present study investigated the possibilities of assessing student creativity based on neural networks approaches for image analysis. The use of psychometric data analysis in the methodology of Latent Class Analysis (LCA) allowed us to obtain data labels to train the neural network without experts' involvement. The high accuracy in network predictions for identifying image creativity suggested large-scale prospects for machine learning to assess complex educational and psychological characteristics.

Keywords: creativity, image analysis, neural networks, educational assessment, psychometrics, machine learning.

DOI 10.14357/20718594210109

References

1. Forum W. E. New vision for education: Unlocking the potential of technology / W. E. Forum, British Columbia Teachers' Federation Vancouver, BC, 2015.
2. Griffin P., Care E. Assessment and teaching of 21st century skills: Methods and approach / P. Griffin, E. Care, Springer, 2014.
3. Lucas B., Claxton G., Spencer E. Progression in student creativity in school: First steps towards new forms of formative assessments 2013.
4. Приказ Министерства образования и науки РФ от 6 октября 2009 г. N 373 «Об утверждении и введении в действие федерального государственного образовательного стандарта начального общего образования».
5. Приказ Министерства образования и науки РФ от 17 декабря 2010 г. N 1897 «Об утверждении федерального государственного образовательного стандарта основного общего образования».
6. Chuang T.-Y., Liu E. Z.-F., Shiu W.-Y. Game-based creativity assessment system: the application of fuzzy theory // *Multimedia Tools and Applications*. 2015. № 21 (74). P. 9141–9155.
7. Sekeroglu B., Dimililer K., Tuncal K. Student performance prediction and classification using machine learning algorithms 2019. P. 7–11.
8. Shao, Z., Li, Y., Wang, X., Zhao, X., Guo, Y. Research on a new automatic generation algorithm of concept map based on text analysis and association rules mining // *Journal of Ambient Intelligence and Humanized Computing*. 2020. № 2 (11). P. 539–551.
9. Liu, O. L., Rios, J. A., Heilman, M., Gerard, L., Linn, M. C. Validation of automated scoring of science assessments // *Journal of Research in Science Teaching*. 2016. № 2 (53). P. 215–233.
10. Nehm R. H., Ha M., Mayfield E. Transforming biology assessment with machine learning: automated scoring of written evolutionary explanations // *Journal of Science Education and Technology*. 2012. № 1 (21). P. 183–196.
11. Leacock C., Chodorow M. C-rater: Automated scoring of short-answer questions // *Computers and the Humanities*. 2003. № 4 (37). P. 389–405.
12. Torrance E. P. Predictive validity of the torrance tests of creative thinking. // *The Journal of creative behavior*. 1972.
13. Van der Walt S. [et al.]. scikit-image: image processing in Python // *PeerJ*. 2014. (2). P. e453.
14. Pei B., Xing W., Lee H.-S. Using automatic image processing to analyze visual artifacts created by students in scientific argumentation // *British Journal of Educational Technology*. 2019. № 6 (50). P. 3391–3404.
15. Kil I. Development and Preliminary Validation of Image-enabled Process Metrics for Assessment of Open Surgery Suturing Skill 2019.
16. Frischknecht, A. C., Kasten, S. J., Hamstra, S. J., Perkins, N. C., Gillespie, R. B., Armstrong, T. J., & Minter, R. M. The objective assessment of experts' and novices' suturing skills using an image analysis program // *Academic Medicine*. 2013. № 2 (88). P. 260–264.
17. Ward T. B. Structured Imagination: the Role of Category Structure in Exemplar Generation // *Cognitive Psychology*. 1994. № 1 (27). P. 1–40.
18. Jankowska D. M., Karwowski M. Measuring creative imagery abilities // *Frontiers in psychology*. 2015. (6). P. 1591.
19. Kim S. H., Vincent L. C., Goncalo J. A. Outside advantage: Can social rejection fuel creative thought? // *Journal of Experimental Psychology: General*. 2013. № 3 (142). P. 605.
20. Slavutskaya E., Abrukov V., Slavutskii L. Neural Network Systems Analysis of the Split-level Psychological Characteristics // *Vestnik Chuvashskogo universiteta*. 2016. №1. Available at: <https://cyberleninka.ru/article/n/neyrosetevoy-sistemnyy-analiz-urovneyh-psihologicheskikh-harakteristik> (accessed December 2, 2020).
21. Uglanova, I., Orel, E., Brun, I. Measuring creativity and critical thinking in primary school // *Psikhologicheskii zhurnal*. 2020. 41. №6. Available at: <http://ras.jes.su/psy/s020595920011124-2-1> (accessed November 28, 2020).
22. Bogoyavlenskaya D.B. *Psihologiya tvorcheskih sposobnostei*. Moscow: Fedorov. 2009. 342 p.
23. Sternberg R. J. *Handbook of creativity* / R. J. Sternberg, Cambridge University Press, 1999.
24. Clark S. L., Muthén B. Relating latent class analysis results to variables not included in the analysis / S. L. Clark, B. Muthén, 2009.
25. Nylund-Gibson K., Choi A. Y. Ten frequently asked questions about latent class analysis. // *Translational Issues in Psychological Science*. 2018. № 4 (4). P. 440.
26. Muthén L. K., Muthén B. O. *Mplus user's guide* / Los Angeles, CA: Muthén & Muthén, 2007.
27. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. Gradient-based learning applied to document recognition // *Proceedings of the IEEE*. 1998. № 11 (86). P. 2278–2324.
28. Krizhevsky A., Sutskever I., Hinton G. E. Imagenet classification with deep convolutional neural networks // *Communications of the ACM*. 2017. № 6 (60). P. 84–90.
29. He, K., Zhang, X., Ren, S., & Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification 2015. P. 1026–1034.
30. *Deep Neural Networks Applications in Handwriting Recognition*. Available at: http://technodocbox.com/3D_Graphics/70716176-Deep-neural-networks-applications-in-handwriting-recognition.html (accessed November 16, 2020).
31. Raghu, M., Zhang, C., Kleinberg, J., Bengio, S. Transfusion: Understanding transfer learning for medical imaging // *Advances in neural information processing systems*. 2019. P. 3347–3357.

32. DENSENET. Available at: https://pytorch.org/hub/pytorch_vision_densenet (accessed November 23, 2020).
33. Von Davier A. A. [et al.]. Computational psychometrics approach to holistic learning and assessment systems // *Frontiers in Education*. 2019. №4. P. 69.

Uglanova I.L. Research interests: psychometrics, measurement in psychology and education. E-mail: iuglanova@hse.ru

Gelver E.S. Senior Business Analyst at Glowbyte Consulting. Research interests: machine learning, computer vision, natural language processing. E-mail: evgelver@gmail.com.

Tarasov S.V. Research of interests: measurement in psychology and education, progress in education. E-mail: svtarasov@hse.ru

Gracheva D.A. Research of interests: psychometrics, comparability of test results. E-mail: dgracheva@hse.ru

Vyrva E. E. Research of interests: test development, measurement in psychology and education. E-mail: evyrva@hse.ru