DOI 10.37882/2223-2966.2025.02.31

СНИЖЕНИЕ ВЛИЯНИЯ ОШИБОК В ОБУЧАЮЩИХ ДАННЫХ ДЛЯ НЕЙРОННЫХ СЕТЕЙ НА ПРИМЕРЕ МЕДИЦИНСКОГО ОБРАЗОВАНИЯ

REDUCING THE IMPACT OF TRAINING DATA ERRORS ON TRAINING OF NEURAL NETWORKS IN MEDICAL EDUCATION

V. Tarasov

Summary. Artificial intelligence systems are actively being introduced into medical education. However, the reliability of AI systems depends on the quality of training data, which can be contaminated with various types of errors. These errors can lead to incorrect training results and have a negative impact on the quality of training and even on the health of patients. The main goal is to study the impact of various types of errors in training data on the training of neural networks used in medical education and to develop methods for eliminating their negative impact based on existing methods and algorithms. This article applies the following methods: classification of error types in training data, analysis of existing algorithms for reducing noise in training data and identification of their limitations with respect to intentional errors, modeling of neural network training and development and application of an improved EM (Expectation-Maximization) algorithm that takes into account the time it takes students to solve problems for a more accurate estimate of the noise distribution in labels. The experiments demonstrate a significant increase in the accuracy of neural network training when using the improved EM algorithm compared to the traditional approach.

Keywords: noise labels, training accuracy, problem solving time, adaptive optimization, feature vector, EM algorithm.

Тарасов Вячеслав Сергеевич
Аспирант, Российский технологический университет, Московский государственный технический университет радиотехники, электроники и автоматики
slavatarasov207@gmail.com

Аннотация. Системы искусственного интеллекта активно внедряются в медицинское образование. Однако, надежность ИИ-систем зависит от качества обучающих данных, которые могут быть зашумлены различными типами ошибок. Эти ошибки могут приводить к неверным результатам обучения и оказывать отрицательное влияние на качество обучения и даже на здоровье пациентов. Основной целью является исследование влияние различных ошибок в обучающих данных на нейронные сети, применяемых в медицинском образовании, и разработка методов устранения их негативного влияния на основе существующих методов и алгоритмов. В данной статье применяются методы: классификации типов ошибок в обучающих данных, анализ существующих алгоритмов для снижения шума в обучающих данных и выявление их ограничений в отношении преднамеренных ошибок, моделирование обучения нейронной сети и разработка и применение усовершенствованного алгоритма EM (Expectation-Maximization), учитывающего время решения задач обучающимися для улучшенной точности оценки распределения шума в метках. Проведенные эксперименты демонстрируют значительное увеличение точности обучения нейронной сети при использовании усовершенствованного алгоритма EM по сравнению с традиционным подходом.

Ключевые слова: шумовые метки, точность обучения, время решения задач, адаптивная оптимизация, Вектор признаков, алгоритм EM.

Введение

В области медицинского образования системы искусственного интеллекта (ИИ-системы) используются для сегментации и классификации деятельности обучающихся в системе массовых открытых онлайн курсов (МООК). В результате начинающейся всеобщей зависимости от ИИ-систем важно они были надежны и на них не влияли сторонние факторы, которые могут привести к ошибке во входных наборах данных и могут серьезно повлиять на пользователей и в дальнейшем, при не качественном обучении при помощи нейронной сети и на пациентов.

В данной работе исследуется влияние и снижение чувствительности к ошибкам обучающих наборов данных при работе с ИИ-системами, а также методы снижения последствия данных ошибок на обучение ИИ. Такое исследование выполняется путем моделирования обучения с применением улучшенного способа отсеивания шума в используемом наборе данных в соответствии с различными значениями, а затем измерения точности системы искусственного интеллекта. В ходе проведенной работы затрагивается вопрос насколько эффективно можно отформатировать данные для этапа обучения.

Цели работы — исследовать эффективность применения улучшенного способа отсеивания шума в наборе данных для искусственного интеллекта, продемонстрировать влияние ошибок обучения на точность нейронной сети.

Описание типов ошибок и их связь с этическими проблемами

Ошибки можно подразделить на различные типы в системах обучающих данных для ИИ. В категории

входят ошибок обучения и ошибок логического вывода. Категории ошибок обучения относятся следующие: (1) невинные / случайные ошибки, (2) преднамеренные ошибки, (3) ошибки типа пропусков и (4) ошибки данных. К случайным ошибкам можно отнести непреднамеренно, например, сделанные человеком, неправильно маркирующим некоторые изображения в тесте, или неправильно помечающим некоторые элементы заданий в обучающем наборе данных. Этот тип ошибок может возникать неизбежно.

Под преднамеренными ошибками можно понимать, намеренно вносимые неправильные ответы, например в тестированиях, такие как повреждение и контаминация данных или преднамеренная неправильная маркировка или вставка неверных, неточных, ненадежных или предвзятых данных, которые могут повлиять на систему искусственного интеллекта, например работать некорректно или иным образом повлиять на результат работы системы искусственного интеллекта по любой причине. Такие преднамеренные ошибки могут способствовать в качестве облегчения обучения для недобросовестных студентов при недобросовестном использовании обучающей системы в которой применяется нейронная сеть, при этом учащийся стремится оставаться в результате тестирования на одном уровне, а не решать сложные задачи. Такие ошибки могут так же возникать в результате скрытого замысла или по различным другим причинам. Вводя такие преднамеренные ошибки, обучающийся может изменить результат работы системы искусственного интеллекта для получения результатов, соответствующих его намерениям.

Ожидается, что после такой неверной маркировки и последующего обучения система искусственного интеллекта на этапе вывода выдаст результаты, которые не смогут предоставить правильный подход к обучению в зависимости от остаточных знаний обучающегося, параметр оценки остаточных знаний не интересующихся обучением студентов может совпасть с людьми, полагающимися на такие результаты, они будут с меньшей эффективностью изучать предоставляемый им обучающий материал [22].

В другом, гипотетическом примере, обучающийся может намеренно вводить ошибки, чтобы вызвать неправильное обучение системы искусственного интеллекта. В таком сценарии, предполагая, что злоумышленник имеет доступ к обучающим данным системы искусственного интеллекта ведь нейронная сеть учится на результатах прошедших ранее обучающихся. Затем обучающийся может снова обучить систему искусственного интеллекта, чтобы подготовить ее к выполнению вывода, который, как ожидается, приведет к результатам, отклоняющим запросы на верный обучающий материал, например показать 100 % знания по всем темам, не изучаемым его курсом или направлением при переподготовке. По мере того, как системы искусственного интеллекта становятся все более повсеместными и на их результаты, со временем, будет полагаться все больше людей, ожидается, что опасность преднамеренных ошибок возрастет.

Ошибки типа пропусков могут возникать по различным причинам, включая, например, отсутствие достаточного разнообразия обучающих данных, небрежность, недостаточную осведомленность и игнорирование сбоя системы искусственного интеллекта [14].

Что касается ошибок в данных, следует относится к ним, как к ошибкам, которые могут быть обнаружены в обучающем наборе данных, которые являются результатом как шум или нестабильности систем глубокого обучения в отношении реконструкции / классификации задач [23].

Некоторые из вышеупомянутых ошибок могут принимать одну или несколько форм. Например, преднамеренная ошибка может быть ошибкой, вызванной вредоносным патчем, или ошибкой, вызванной неправильной маркировкой, и так далее.

Следует отметить, что каждый из вышеупомянутых ошибок обучения и ошибок вывода не обязательно являются исчерпывающими, также могут возникать другие типы ошибок.

Анализ алгоритмов для снижения шума в обучающем наборе данных

Зашумленные обучающие данные и, в частности, ошибки маркировки были предметом многих исследований. Бродли и Фридл используют набор алгоритмов обучения для создания классификаторов, которые служат в качестве фильтров шума. Исследователи оценивают единый алгоритм, способы фильтрации голосования большинством голосов и консенсуса и показывают, что фильтрация значительно повышает точность классификации по уровням шума до 30%. [6]

Беккер и Голдбергер предлагают алгоритм обучения нейронных сетей, основанный исключительно на зашумленных данных. Они вводят дополнительный слой шума, предполагая, что наблюдаемые метки создаются из истинных меток путем прохождения через зашумленный канал, и показывают, что они могут изучать распределение шума из зашумленных данных без использования каких-либо чистых данных. Голдбергер и Бен-Реувен моделируют шум с помощью дополнительного слоя softmax, который соединяет правильные метки с зашумленными. [6]

Дгани, Гринспен и Голдбергерт используют зашумленный канал в своей стратегии обучения нейронных сетей, которая основана на ненадежных комментариях человека.

Гуан и др. представляют подход к идентификации и устранению неправильно помеченных обучающих экземпляров путем использования размеченых экземпляров для облегчения обнаружения неправильно помеченных обучающих экземпляров, и Хошгофтаар и др.

Упомянутые выше работы имеют дело с зашумленными данными, которые обычно связаны с процессом сбора данных. Как таковые, работы, вероятно, могут подходить для типа ошибок обучения, которые называем ошибками данных, но не обязательно для намеренно сделанных ошибок категории ошибок обучения, которые, как ожидается, не будут смоделированы как шум.

Кроме того, в отношении дисбаланса классов, когда он создается намеренно, то есть когда он является частью преднамеренных ошибок, недобросовестный пользователь может решить усилить дисбаланс классов или изменить важные примеры классов, которые составляют основу для пере дискретизации и занижения дискретизации, чтобы увеличить количество ошибок в экземплярах данных, и, следовательно, такой дисбаланс классов не может рассматриваться методами например пересэмплирования или недосэмплирования выборки данных как дисбаланс классов в собранных данных.

Кроме того, также можно предположить, что в отношении намеренно допущенных ошибок злоумышленник может вставить их после того, как фильтры и другие алгоритмы, предназначенные для повышения точности классификации и баланса классов, уже были использованы. Соответственно, упомянутые выше работы не решают этических и функциональных проблем, связанных с некоторыми типами ошибок обучения, и необходимы дальнейшие действия, процедуры и инструменты, чтобы помочь справиться с такими этическими и функциональными проблемами.

Анализ алгоритма «Expectation-Maximization» для обучения нейронной сети с шумовыми метками

Так для избавления от шума в виде ошибок для приближения нейронной сети к полноценной работе можно использовать способ из работы Беккер, Дж., Голдбергер. Этот подход основан на обучении нейронной сети, использовавшим исключительно зашумлённые данные, что является интересным и перспективным решением для ситуаций, когда доступ к чистым данным ограничен. [6]

Для обучения мульти классовых нейросетевых мягкой классификации p ($y = i|x; w$), где x — вектор признаков, а w — набор параметров сети и i является членом множества классов {1, ..., k}. Также можно предположить, что в процессе обучения не можем напрямую наблюдать правильную метку y. Вместо этого у есть доступ только к ее зашумленной версии, обозначенной z. В подходе генерация шума моделируется параметром θ (i, j) = p ($z = j|y = i$) [6]. Распределение шума неизвестно, и хотим изучить его как часть фазы обучения. Вероятность наблюдения зашумленной метки z при заданном векторе признаков x равна:

$$p(z = j \mid x;w,\theta) = \sum_{i=1}^{k} p(z = j \mid y = i;\theta)p(y = i \mid x;w)$$

где k — количество классов. Модель проиллюстрирована рис. 1.

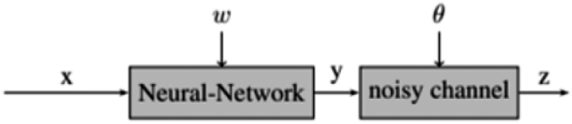

Рис. 1. Модель зашумленной метки z

Предположим, что классификатор нейронной сети, который используются, основан на нелинейных промежуточных слоях, за которыми следует выходной слой softmax, используемый для мягкой классификации. Обозначим нелинейную функцию, применяемую к входу x, как $h = h(x)$, а обозначим слой softmax как:

$$p\left(y = i \;\; x;w\right) = \frac{\exp\left(u_i^T h\right)}{\sum_{j=1}^{k} exp\left(u_i^T h\right)}$$

u_1, ..., u_k являются параметрами мягкого максимума, которые являются подмножеством всего набора параметров сети w.

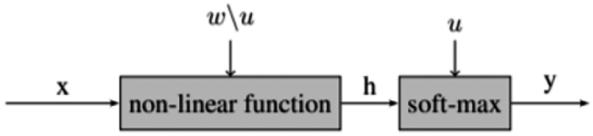

Рис. 2. Модель нелинейной функции

На этапе обучения даны n векторов признаков x, ..., x_n с соответствующими ненадежными метками z, ..., z_n которые рассматриваются как шумовые версии правильных скрытых меток y, ..., y_n. Логарифмическое правдоподобие параметров модели равно:

$$L\left(\omega,\theta\right) = \sum_{t=1}^{n} \log\left(\sum_{i=1}^{k} p(z_t|y_t = i;\theta)p(y_t = i \mid x_t;\omega)\right)$$

Цель шага E состоит в том, чтобы на основе данных обучения найти как распределение шума θ, так и параметры нейронной сети w, которые максимизируют функцию правдоподобия. Поскольку случайные величины y ..., y_n скрыты, применяем алгоритм «Expectation-Maximization» далее (EM) для поиска набора параметров максимального правдоподобия.

Применение алгоритма нейронной сети с шумовыми метками позволит получить чистый результат в подстраивании при обучении нейронной сети, которая поможет в обучении.

Вход: Точки данных $x, ..., x_n \in Rd$ с соответствующими шумными метками $z, ..., z_n \in \{1, ..., k\}$.

Выход: Параметры нейронной сети w и шумовые параметры θ. Алгоритм EM выполняет итерации между двумя шагами:

Шаг E: Оценка истинных меток на основе текущих значений параметров:

$$C_{ti} = p(y_t = i | x_t, z_t; \omega, \theta)$$

M-шаг: обновление параметра шума θ:

$$\theta(i,j) = \frac{\sum_t c_{ti} 1_{\{z_t = j\}}}{\sum_t c_{ti}}$$

и обучить NN для нахождения w, которые максимизируют следующую функцию:

$$L(\omega) = \sum_{t=1}^{n} \sum_{i=1}^{k} c_{ti} \log p(y_t = i \mid x_t; \omega)$$

Конечно, нет необходимости полностью обучать модель NN на каждой итерации EM. Например, можно использовать стандартные методы для обучения нейронной сети и обновлять параметр шума θ после нескольких проходов по обучающему набору. Алгоритм EM является жадной процедурой оптимизации и, чувствителен к начальной точке. Следовательно, хорошая инициализация параметров модели важна для достижения хороших результатов.

Повышение точности обучения нейронной сети путем усовершенствования алгоритма EM

Основной алгоритм EM для оценки распределения шума в метках обучающих данных опирается на вероятность принадлежности точки данных к каждому классу. Однако, дополнительная информация о времени, затраченном на выполнение задачи, может помочь улучшить оценку шума. Точная оценка параметра шума $\theta(i,j)$ способствует правильному обучению нейронной сети, которая будет лучше учитывать фактор шума в обучающих данных. Стандартные отклонения σ_j необходимо оценивать на основе исторических данных о решении задач, вследствие чего изначальные данные должны быть достоверны. Этот подход предполагает наличие данных о времени решения задач обучающимся. В знаменателе — это взвешенная сумма вероятностей C_{ti}. Веса $\exp(-(t_i - t_j)^2 / (2\sigma_i^2))$ определяются временем решения: чем ближе время решения к среднему времени решения для j-го класса, тем больше вес. В числителе — то же самое, но умноженное на индикаторную функцию $1\{z_t = j\}$. Таким образом, учитывается, что ошибки, сделанные за время, близкое к среднему времени решения для j-го класса, с большей вероятностью могут быть случайными. Так применение такого метода отраженно в следующей функции:

$$\theta(i,j) = \frac{\sum_t c_{ti} 1_{\{z_t = j\}} * exp - \left((t_i - t_j)^2 / 2\sigma_j^2 \right)}{\sum_t c_{ti} * exp - \left((t_i - t_j)^2 / 2\sigma_j^2 \right)}$$

t_j — среднее время решения задач для j-го класса.

σ_j — стандартное отклонение времени решения задач для j-го класса.

Преимуществом представленного выше метода может быть точная оценка шума в метках. Эта формула учитывает дополнительную информацию о времени решения, что позволяет точно оценивать вероятность шума в метках. Следует учитывать, что стандартные отклонения σ_j необходимо оценивать на основе исторических данных о решении задач. Точная оценка параметра шума может повысить качество обучения нейронной сети, так как модель будет лучше учитывать фактор шума в данных.

Эксперимент

В этом разделе оцениваем надежность глубокого обучения к обучающим данным с шумными метками с явным моделированием шума и без него. В наших экспериментах использовали два набора данных с введенным шумом меток. В базу данных входят следующие значения, где x1 — это общее количество времени, проведенное в системе, x2 — клики по конспекту лекции, x3 — оценка за задание по конспекту лекции, x4 — оценка за проверочный тест по теме «Введение. МЧС», x5 — клики по заданиям лабораторной работы № 1, x6 — клики по заданиям лабораторной работы № 1, x7 — клики по конспекту лекции по теме «Методы эвакуации МЧС», x8 — оценка за задание по конспекту лекции по теме «Введение», x9 — оценка за проверочный тест по теме «Введение», x10 — клики по заданиям лабораторной работы № 2, x11 — клики по конспекту лекции «Базовые операции», A — средний бал за тестирование.

При проведении обучения нейронной сети без предварительной подготовки входных данных предполагаемых обучающихся с показателями, которые указаны выше, получаем точности в зависимости от эпох на рисунке 1.

Так переходя к использованию EM-алгоритма код, реализует для оценки распределения шума в метках. Идет расчёт вероятность принадлежности каждой точки дан-

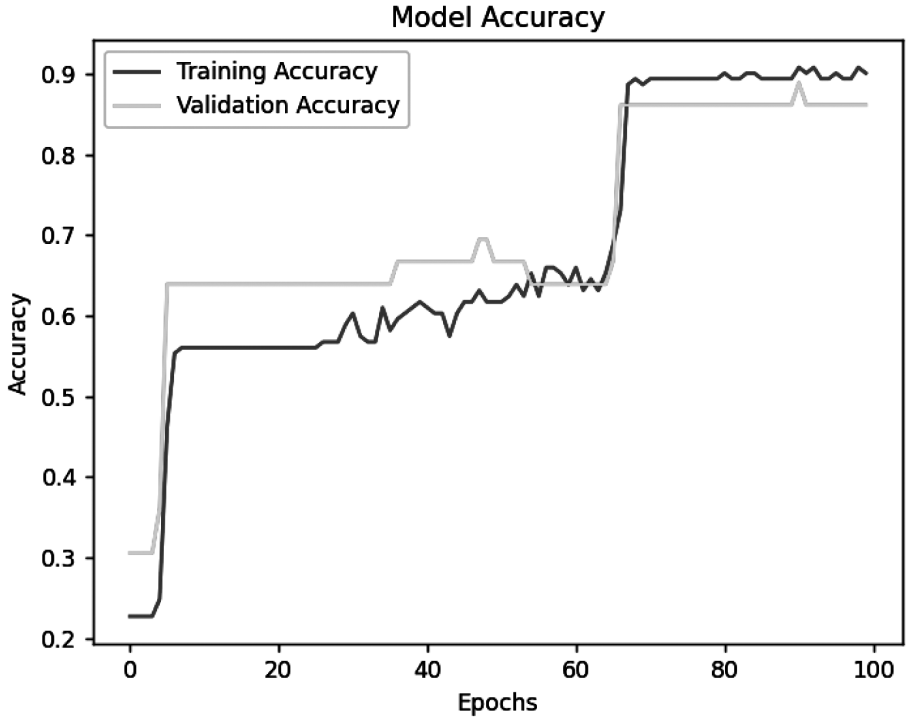

Рис. 3. Точность обучения нейронной сети при первоначальных данных

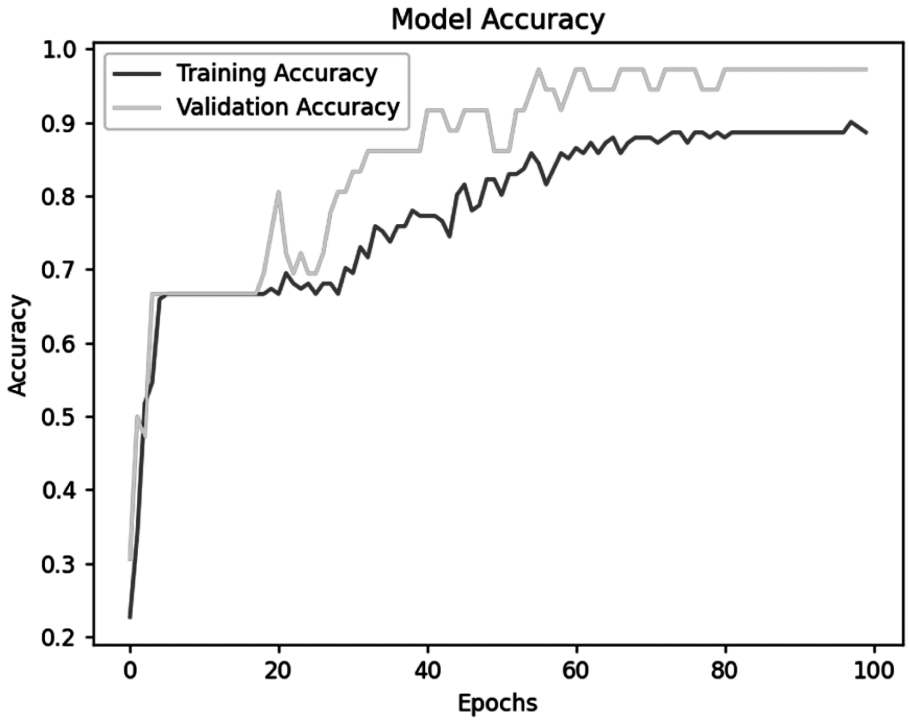

Рис. 4. Точность обучения нейронной сети при первоначальных данных

ных к каждому из классов, учитывая текущие параметры модели. Обновляются параметры шума (матрица theta) на основе вычисленных вероятностей.

Если применить алгоритм EM к нейронной сети оценки остаточных знаний, которая использовалась ранее, видим, что точность обучения повышается значительно и плавно при обучении в 100 эпох. Это говорит о том, что алгоритм EM способен эффективно справляться с шумом в данных, присутствующем в результатах обучения нейронной сети, которая оценивает остаточные знания обучающихся. Важно отметить, что применение алгоритма EM приводит к стабильному и предсказуемому росту точности обучения продемонстрированное на рисунке 4.

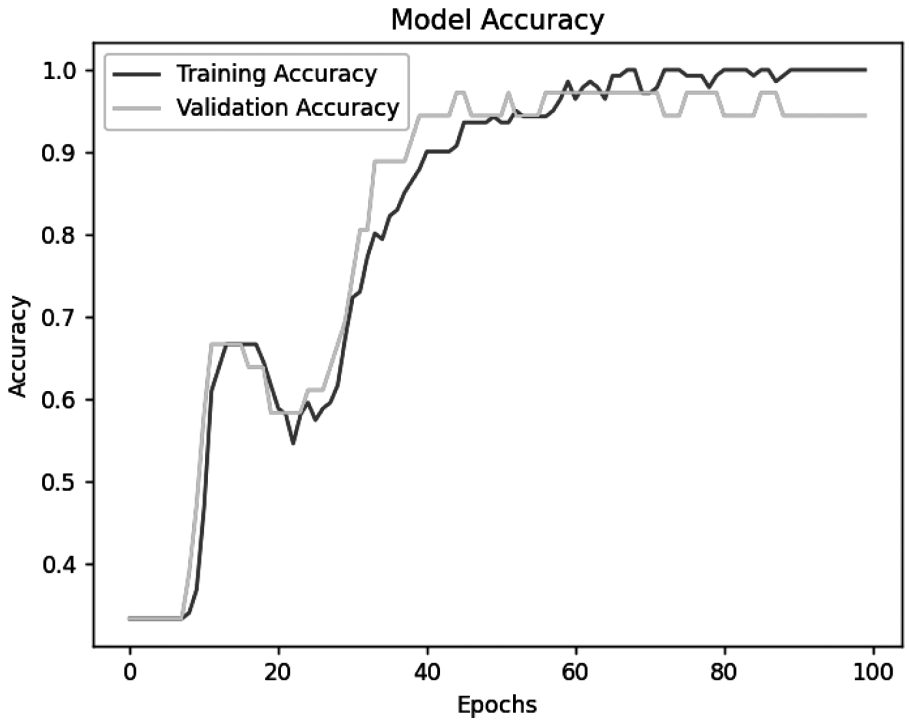

Рис. 5. Точность обучения нейронной сети при использования улучшенного алгоритма EM

Это улучшение достигается за счет того, что алгоритм EM учитывает вероятность принадлежности каждой точки данных к каждому из классов, учитывая текущие параметры модели и корректирует параметры шума на основе вычисленных вероятностей. Таким образом, алгоритм EM позволяет нейронной сети точно учитывать шум в данных, что приводит к точным результатам оценки остаточных знаний обучающихся.

При использовании в ходе эксперимента улучшенный алгоритм можем наблюдать на рисунке 3 что точность обучения нейронной сети значительно повышается по сравнению с использованием традиционного алгоритма EM. Рисунок 3 демонстрирует плавный и устойчивый рост точности обучения в течение 100 эпох, что свидетельствует о эффективном и надежном обучении нейронной сети при учете времени решения задач обучающимися.

Так с применением обновлённого алгоритма процесс и точность обучения поднимается гораздо выше при использовании входных данных, которые были использованы прежде, но уже с отсеиванием ненадежных параметров особенно, учитывая время выполнения обучающимся работы. Так применение нового алгоритма способствует в дальнейшем производить лучшие оценочные данные нежели с использованием старых параметров.

Заключение

При изучении существующих алгоритмов для снижения шума в обучающих данных был выбран один наиболее подходящий для улучшения и использования с нейронной сетью по оценке остаточных знаний для снижения зашумленности данных при обучении ИИ. В качестве решения предлагается новый подход, основанный на алгоритме EM, который позволяет оценивать распределение шума в метках и корректировать обучение нейронной сети. Эксперименты показали, что применение модифицированного алгоритма EM значительно повышает точность обучения нейронной сети по сравнению с использованием исходных данных.

Ошибки в обучающих данных представляют серьезную угрозу для надежности ИИ— систем, особенно в медицинской сфере. Существующие методы борьбы с шумом не всегда эффективны для устранения преднамеренных ошибок, так как в ходе данной работы и эксперимента существующие алгоритмы обладают потенциалом к улучшению своих показателей. Предложенные улучшения алгоритма EM является перспективным решением для повышения устойчивости ИИ-систем к ошибкам обучения вызванными невинными действиями, случайными ошибками или иными действиями.

ЛИТЕРАТУРА

1. Frenay and A. Kaban. A comprehensive introduction to label noise. In ´ European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), 2014.
2. G.E. Hinton, O. Vinyals, and J. Dean. Distilling the knowledge in a neural network. In NIPS Deep Learning and Representation Learning Workshop, 2014.
3. D. Kingma and J. Ba. Adam: A method for stochastic optimization. arXiv preprintarXiv:1412.6980, 2014.
4. S. Reed, H. Lee, D. Anguelov, C. Szegedy, D. Erhan, and A. Rabinovich. Training deep neural networks on noisy labels with bootstrapping. In arXiv preprint arXiv:1412.6596, 2014.
5. S. Sukhbaatar and R. Fergus. Learning from noisy labels with deep neural networks. In arXiv preprint arXiv:1406.2080, 2014.
6. Bekker and J. Goldberger. Training deep neural networks based on unreliable labels. In IEEE Int.l Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 2682 2686, 2016.
7. Brady A.P.: Error and discrepancy in radiology: inevitable or avoidable? Insights Imaging 8, 171–182 (2017)
8. E. Beigman and B.B. Klebanov. Learning with annotation noise. In ACL-IJCNLP, 2009. 8 Published as a conference paper at ICLR 2017
9. Mintz Y., Brodie R.: Introduction to artificial intelligence in medicine. Min. Inv. Ther. All. Tech. 28, 2–73 (2019)
10. Bredt S.: Artificial Intelligence (AI) in the financial sector—Potential and public strategies. Front. AI. 2, 3 (2019)
11. Kirtan J., Aalap D., Poojan P., Manan S.: A comprehensive review on automation in agriculture using artificial intelligence. AI. Agr. 2, 1–2 (2019)
12. Finlayson S.G., Bowers J.D., Ito J., Zitrain J.L., Beam A.L., Kohane I.S.: Adversarial attacks on medical machine learning. Science 1, 1287–1289 (2019)
13. Challen R., Denny J., Pitt M., et al.: Artificial intelligence, bias and clinical safety. BMJ Qual. Saf. 28, 231–237 (2019)
14. Geis J.R., Brady A.P., Wu C.C., et al.: Ethics of artificial intelligence in radiology: summary of the joint European and North American Multisociety Statement. Radiology 293(2), 436–440 (2019)
15. Zannettou S., Baumgartner J., Finkelstein J., Goldenberg A.: Weaponized information outbreak: a case study on COVID-19. Bioweapon Myths and the Asian Conspiracy Meme (2019)
16. Kaissis G.A., Makowski M.R., Rückert D., et al.: Secure, privacy-preserving, and federated machine learning in medical imaging. Nat. Mach. Intell. 2, 305–311 (2020)
17. Qayyum A., Qadir J., Bilal M., Al-Fuqaha A.: Secure and robust machine learning for healthcare: a survey. Preprint at https://arxiv.org/abs/2001.08103 (2020)
18. Strickland E.: Healthcare algorithms show racial bias. IEEE Spect. 8, 6–7 (2020)
19. D'Antonoli T.A.: The ethical considerations for artificial intelligence: an overview of the current radiology landscape. Diagn. Interv. Radiol. 26, 504–511 (2020)
20. Larrazabal A.J., Nieto N., Peterson V., Milone D.H., Ferrante E.: Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. PNAS 117(23), 12592–12594 (2020)
21. Castelvecchi D.: Is facial recognition too biased to be let loose? Nature 587, 347–349 (2020)
22. Antun V., Renna F., Poon C., Adcock B., Hansen A.C.: On instabilities of deep learning in image reconstruction and the potential costs of AI. Proc. Nat. Aca. Sci. 117(48), 30088–30095 (2020)
23. Hutson M.: It's too easy to hide bias in deep-learning systems. IEEE Spect. 1, 2–19 (2021)

Журнал «Современная наука: актуальные проблемы теории и практики»