

1. Введение

Современные задачи компьютерного зрения часто сталкиваются с проблемой работы с реальными данными, которые могут содержать ошибки в разметке, шумы и неопределенности. Такие данные представляют собой серьезный вызов для разработки устойчивых и надежных моделей машинного обучения, способных эффективно функционировать в условиях неидеальных входных данных. Одним из ключевых аспектов в этой области является разработка методов, которые не только минимизируют ошибки классификации, но и способны оценивать неопределенность предсказаний, что особенно важно для приложений, где ошибочные решения могут иметь серьезные последствия.

В данной работе рассматривается задача классификации на наборе данных CIFAR-10N, который содержит зашумленные метки, что делает его идеальным полигоном для изучения методов работы с несовершенными данными. Основная цель исследования — сравнение различных функций потерь, которые могут быть использованы для обучения моделей, способных не только классифицировать данные, но и оценивать неопределенность своих предсказаний. В частности, рассматриваются три типа функций потерь: стандартная кросс-энтропийная функция потерь (CE-loss), пользовательская функция потерь (B-loss), которая оценивает неопределенность предсказаний, и гетероскедастическая регрессионная функция потерь (N-loss), основанная на многомерном нормальном распределении.

2. Связанные работы

В работе Han et al. [8] предложен метод Co-teaching, который предназначен для обучения глубоких нейронных сетей на данных с зашумленными метками. Основная идея заключается в обучении двух моделей параллельно, где каждая модель выбирает "чистые" данные для обучения другой модели. Это позволяет уменьшить влияние зашумленных меток на процесс обучения.

Co-teaching может быть использован для улучшения устойчивости моделей к зашумленным данным в CIFAR-10N. В частности, можно адаптировать этот метод для обучения ансамблей моделей, что может повысить точность и устойчивость предсказаний.

Collier et al. [3] предложили метод для обучения глубоких классификаторов в условиях гетероскедастического шума в метках. Они используют настройку температуры softmax для управления компромиссом между смещением и дисперсией, что позволяет модели лучше справляться с зашумленными данными.

Метод может быть использован для улучшения устойчивости моделей к зашумленным данным в CIFAR-10N. В частности, можно интегрировать настройку температуры softmax в предложенные модели (B-loss и N-loss) для улучшения их производительности.

Wei et al. [23] исследовали влияние сглаживания меток (label smoothing) на обучение моделей с зашумленными метками. Они показали, что сглаживание меток может улучшить устойчивость моделей к шуму, особенно когда уровень шума не слишком высок.

Сглаживание меток может быть использовано в данной работе для улучшения устойчивости моделей к зашумленным данным. В частности, можно применить этот метод к моделям, обученным с использованием B-loss и N-loss, чтобы оценить его влияние на точность и устойчивость.

Xia et al. [25] предложили метод выборки с учетом неопределенности (uncertainty-aware sample selection), который позволяет модели выбирать наиболее информативные данные для обучения, учитывая неопределенность предсказаний. Это особенно полезно для работы с зашумленными данными, так как модель может избегать примеров с высокой неопределенностью.

Метод Xia et al. может быть использован для улучшения качества обучения моделей на CIFAR-10N. В частности, можно интегрировать оценку неопределенности в процесс обучения моделей с использованием B-loss и N-loss, чтобы улучшить их устойчивость к зашумленным данным.

4 Результаты и обсуждение.

Набор данных CIFAR-10N [24] был разбит на обучающий, валидационный и тестовый наборы в количестве [45000, 5000, 10000] образцов соответственно. Модели обучались с помощью 9-слойной светочной нейронной сети [8, 25], Сеть имеет 4,4 миллиона параметров, которые случайным образом инициализируются в процессе обучения. Архитектура CNN и большинство настроек соответствуют экспериментам Xia et al. [25] с небольшими изменениями: модели обучались в течение 20 эпох (200 в оригинальной статье) с использованием оптимизатора Adam с импульсом 0.9 и размером батча 128, а также с постоянной скоростью обучения 0.001 (в оригинальной статье начальная скорость обучения линейно уменьшалась до нуля, начиная с 80-й эпохи); образцы изображений были преобразованы в тензоры и нормализованы со средними значениями [0,491, 0,482, 0,447] и стандартными отклонениями [0,247, 0,243, 0,261].

Некоторые экспериментальные настройки могут отличаться от тех, что описаны в работе Xia et al. [25]: применялся метод сборки моделей; к образцам во всех наборах в качестве преобразования применялся кадр случайного размера с диапазоном масштабов [0,8, 0,1] и соотношением сторон [0,9, 1,1], поэтому выборка тестовых наборов осуществлялась с помощью дополнения тестового времени кадром случайного размера; для вывода использовались модели с наименьшими потерями при валидации.

Все эксперименты проводились семь раз со случайными семенами [42, 0, 17, 9, 3, 16, 2]. Затем были получены среднее и стандартное отклонение экспериментальных результатов. Многочисленные прогнозы, полученные в результате объединения моделей, позволили вычислить окончательные прогнозы с помощью голосования по большинству голосов.

В качестве метрики использовалась Ассигасу [12]. Полученные результаты не сравнивались с современными.

Таблица 1: Точность (%) моделей, обученных на наборе данных CIFAR-10N с чистыми и зашумленными метками, а также со сглаживанием меток (LS). Каждая модель обучалась с использованием семи различных семян инициализации веса в течение 20 эпох, а затем собиралась в ансамбль. Средняя тестовая точность отдельных моделей сравнивается с точностью ансамбля с использованием мажоритарного голосования (EMV). Лучшие результаты выделены жирным шрифтом. Тип использованного шума random_label3.

Метод	Архитектура	Кол-во параметров	LS	Точность (одиночная модель / EMV) %	
				Чистые данные	Шумные данные
CE Loss	CNN	4.4M	0.1	86.67±0.32 / 89.19	79.49±0.56 / 82.56
B Loss				84.47±0.25 / 88.19	80.47±0.31 / 84.22
N Loss				85.94±0.39 / 88.92	81.38±0.24 / 84.69

5 Этика

Ниже приведен отчет по использованию LLM для конкретных задач в работе

1. Сжатый пересказ источников для оценки их полезности
2. В настоящей работе при разработке кода LLM не применялся
3. LMM использовалась для перевода текста
4. Работа представляла в большей степени описанную методику, поэтому для генерации идеи LLM не использовалась.

Так как в большей степени работа представляла собой проведение эксперимента и настройку платформы для обучения, считаю вклад языковых моделей косметическим и несущественным

6. Заключение

На чистых данных лучше всего себя проявляет CE loss, что, впрочем, неудивительно, иначе не применялся бы он так часто. Однако из-за отсутствия механизма уверенности, при зашумленности данных дает самую низкую точность. Одновременно с этим исследовалось две loss функции, оценивающие уверенность и дисперсию модели. В задачах с зашумленными данным они справились лучше. Метрики показывают, что для описанной задачи более устойчивый к шумам оказался N loss. Это происходит благодаря явному моделированию дисперсии предсказаний, в процессе обучения модель может присвоить шумным данным большую дисперсию, что будет меньше влиять на обучение.