

Университет ИТМО, факультет программной инженерии и компьютерной техники
Двухнедельная отчётная работа по «Информатике»: аннотация к статье

Дата прошедшей лекции	Номер прошедшей лекции	Название статьи/главы книги/видеолекции	Дата публикации (не старше 2022 года)	Размер статьи (от 400 слов)	Дата сдачи
10.09.2025	1	О фундаментальных пределах сжатия данных на основе обучения	21.12.2023	~18000	24.09.2025

Выполнил(а) Махфудх Ахмед Айнин, № группы P3132, оценка _____
Фамилия И.О. студента не заполнять

Прямая полная ссылка на источник или сокращённая ссылка (bit.ly, tr.im и т.п.)

<https://arxiv.org/abs/2312.13847>

Теги, ключевые слова или словосочетания (минимум три слова)

Сжатие данных , Теория информации , Теория скорости-искажения , Машинное обучение , Глубокие генеративные модели

Перечень фактов, упомянутых в статье (минимум четыре пункта)

1. В статье представлена теоретическая основа для анализа фундаментальных пределов сжатия данных, выполняемого моделями на основе обучения .
2. Показано, что для источников с неизвестными распределениями компрессоры на основе обучения могут достигать оптимальной производительности, соответствующей границам классической теории информации.
2. Авторы выводят асимптотическую среднюю производительность по скорости-искажению для компрессоров на основе обучения, связывая ее с внутренними свойствами данных.
3. Ключевой вывод заключается в том, что глубокие генеративные модели могут эффективно изучать сложные распределения данных, что позволяет достигать скоростей сжатия, приближающихся к теоретическим пределам Шеннона .
4. Исследование изучает концепцию «неявного» сжатия, присущую процессу обучения различных архитектур машинного обучения.

Позитивные следствия и/или достоинства описанной в статье технологии (минимум три пункта)

1. Это исследование обеспечивает теоретическое обоснование высокой производительности, наблюдаемой в современных алгоритмах сжатия на основе обучения .
2. Понимание этих фундаментальных пределов может направить разработку еще более эффективных и адаптивных методов сжатия данных .
3. Оно объединяет классическую теорию информации с современным машинным обучением, открывая новые направления для исследований в обеих областях.

Негативные следствия и/или недостатки описанной в статье технологии (минимум три пункта)

1. Достижение этих теоретических пределов на практике часто требует значительных вычислительных ресурсов для обучения сложных моделей машинного обучения.
2. Аспект «обучения» означает, что производительность сильно зависит от качества и количества обучающих данных, которые не всегда могут быть доступны.
3. Представленные теоретические модели могут не полностью охватывать все практические накладные расходы и сложности реальных систем сжатия на основе обучения.

Ваши замечания, пожелания преподавателю или анекдот о программистах¹

Данная статья углубляет понимание пересечения классической теории информации и современных методов машинного обучения, что было очень интересно изучить.

¹ Наличие этой графы не влияет на оценку