1. Добрый день, меня зовут Айрапетов Тигран, тема моей работы – «Символьная регрессия для описания контурных изображений на плоскости». Научный руководитель – Миков Александр Иванович.
2. Актуальность темы обусловлена развитием технологий анализа изображений и, как следствие, необходимостью создания новых подходов к их обработке и анализу.
3. В работе были изучены аспекты предметной области и рассмотрены некоторые методы символьной регрессии.
4. Контурные изображения представляют собой вид графических данных, внимание в которых акцентируется на границах объектов.
5. Символьная регрессия – метод аппроксимации, заключающийся в подборе математического выражения, наилучшим образом отражающего исходные данные.
6. В работе были рассмотрены следующие подходы. Генетическое программирование – подход, вдохновленный эволюционными процессами, где выражения оптимизируются через отбор, скрещивание и мутации. KAN – это нейросеть, основанная на теореме Колмогорова-Арнольда. Ее успешная реализация появилась относительно недавно, а основное отличие от перцептрона – вместо весов мы «обучаем» функции. PySR – оптимизированная для сложных вычислений реализация символьной регрессии основанная на эволюционных алгоритмах совместно с глубоким обучением.
7. При применении символьной регрессии для описания контурных изображений мы можем столкнуться с некоторыми проблемами. Одна из проблем: необходимость сегментировать контуры на подходящие для аппроксимации части. Также, в общем случае, формулу не для любого изображения мы сможем найти за приемлемое время.
8. На слайде можно увидеть разбиение некоторого контурного изображения на контуры. В данном случае это 4 контура: два глаза, рот и все тело. Возникает вопрос: как аппроксимировать такой сложный контур? Можно предположить некоторые методы разделения данных, например кластеризацию. На рисунках вы можете видеть результаты. В одном случае точки разделились довольно удачно, однако на другом примере мы не можем выделить четкого правила разделения. Для решения такого рода проблемы можно использовать нейросети для предобработки и разделения исходных данных.
9. Далее были проведены эксперименты с моделями. На промежутке от -1 до 1 при на 100 точках и 10 поколениях была обучена модель генетического программирования на данной зависимости. Можно отметить схожесть в графиках, однако полученное выражение довольно громоздкое и вряд ли отразит зависимость на других промежутках
10. Но если передать данные на промежутке от -10 до 10 и увеличив обучающую выборку до 1000 можно увидеть, что модель может неплохо справляться при правильной настройке параметров. Однако, итоговое выражение все еще далеко до идеала.
11. На слайде можно увидеть код и пример работы KAN на 10 итерациях на промежутке от -1 до 1. Ошибка была довольно небольшой, однако увеличение числа эпох или обучающей выборки не дало лучшего эффекта. Однако при изменении архитектуры нейросети, она может лучшим образом подстроиться под исходные данные.
12. Далее представлены результаты тестирования PySR. Всего на 1 итерации была получена формула параболы, на 10 итерациях неплохо аппроксимирована более сложная зависимость. Тем не менее, нельзя гарантировать стабильный результат.
13. Модели проявили себя по-разному, но можно выделить общие черты: при аппроксимации сложных зависимостей важно правильно настроить параметры и выделить значительное время. Также важным моментом является уточнение природы исходных данных, так как это может упростить задачу предобработки изображений.
14. Спасибо за внимание