

ИСПОЛЬЗОВАНИЕ ВИЗУАЛЬНЫХ ТРАНСФОРМЕРОВ ДЛЯ ОПРЕДЕЛЕНИЯ ПРОЦЕНТА ЗАСОРА В МЕТАЛЛОЛОМЕ В УСЛОВИЯХ УПЛ ЛПК

Сторонкин Д.А. (ИТМО)

Научный руководитель – Дзюб И.С. (ОМК-ИТ)

Введение. В современном мире металлолом является одним из основных источников вторичного металла, что делает его неотъемлемой частью промышленного производства. Ключевой аспектом обработки и переработки металлолома является определение степени его засора, что влияет на качество получаемого вторичного сырья и стоимость процесса переработки. Традиционные методы определения засора являются трудоемкими, длительными и подвержены ошибкам, так же сильно зависят от экспертного мнения [1]. Поэтому внедрение модели машинного обучения для определения процента засора в металлоломе на участках приемки становится актуальной и перспективной задачей. Данная работа направлена на описание и анализ такой модели, которая может значительно усовершенствовать процесс контроля качества принимаемого металлолома, уменьшить затраты и повысить эффективность всего производственного цикла.

Основная часть. Определение процента засора является задачей регрессии. Регрессия относится к классу задач обучения с учителем, когда по заданному набору признаков наблюдаемого объекта необходимо спрогнозировать некоторую целевую переменную. Набор признаков – трехканальная матрица изображения в цветовом пространстве RGB, целевая переменная – процент засора. Обучение заключается в минимизации $L(\Theta)$ путём коррекции параметров Θ с использованием алгоритмов оптимизации, таких как стохастический градиентный спуск. В процессе обучения мы минимизируем разницу между предсказанным значением и истинным значением из данных обучения, используя функцию потерь - среднюю абсолютную ошибку. Swin Transformer [2] является моделью глубокого обучения, которая была представлена командой исследователей из Microsoft Research. Swin Transformer сочетает в себе идеи механизма внимания [3], широко используемых для обработки естественного языка, и адаптирует их для визуального распознавания. В отличие от традиционных архитектур на основе сверточных нейронных сетей, Swin Transformer полагается исключительно на механизмы самовнимания из архитектуры трансформеров для обработки изображений. В базовом случае данная модель решает задачу классификации изображения и на выходе из блоков связанных непосредственно со Swin Transformer стоит многослойный перцептрон с несколькими выходами, количество выходов было уменьшено до одного.

Выводы. После обучения модели, метрики качества на тестовом наборе данных составили: средняя абсолютная ошибка составляет 0.37, коэффициент детерминации – 0.55. При увеличении данных, которые будут участвовать в обучении – метрики качества будут расти. Целевые метрики качества, которые необходимо достигнуть: средняя абсолютная ошибка составляет 0.3, коэффициент детерминации – 0.6.

Список использованных источников:

1. ГОСТ 2787-2019 Межгосударственный стандарт. Металлы черные вторичные. Общие технические условия.

2. Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, Baining Guo. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. 2021.
3. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. 2017.