

ля (АЦП). Преобразование сигнала в цифровую форму необходимо для более простой линейаризации сигнала датчика и калибровки. Негативным фактором типовых структур, используемых в таких системах АЦП, является необходимость делать выбор между обеспечением разрешающей способностью преобразования и быстродействием (частотой обновления данных на выходе). Поэтому более рациональным является использование усовершенствованного АЦП, позволяющего осуществить прямое подключение емкостного датчика к преобразователю и обеспечивающего высокую разрешающую способность, точность и линейность характеристик.

Разрабатываемый экспериментальный вариант исследования качества машинных масел емкостным способом может быть использован в лекционном курсе дисциплины кафедры «Электроника и измерительная техника» вместе со специально подготовленным методическим обеспечением.

УДК 004.891

ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ЗАДАЧ КЛАССИФИКАЦИИ ТОВАРОВ

**Иваненков Д.А.¹, к.т.н., Кузнецов А.А.², д.т.н., проф.,
Акиндинова Н.С.², к.т.н., доц.**

¹Маркетинговое агентство «50 текс» (резидент НПП ВГТУ),
г. Витебск, Республика Беларусь

²Витебский государственный технологический университет,
г. Витебск, Республика Беларусь

Постоянно увеличивающийся объём данных вызывает необходимость разработки и использования новых, высокопроизводительных алгоритмов автоматической обработки информации. В современном онлайн-пространстве одной из самых востребованных услуг является автоматическая классификация текстовых данных. В данной работе рассматривается способ классификации товарных предложений по разделам на сайте крупного онлайн-агрегатора при помощи нейронных сетей.

Онлайн-площадка публикует информацию о товарах и ценах различных продавцов. Суть проблемы заключается в несоответствии категорий товаров у продавцов и агрегатора. Простое соотношение категорий между сторонами процесса провести невозможно вследствие разностного подхода к классификации.

Сформулируем задачу исследования: сопоставить объекты из входящего множества с заранее известным множеством классов (разделов) на сайте.

На начальном этапе задача решалась с использованием полнотекстового поиска системы управления базами данных MySQL. Fulltext-поиск позволил значительно уменьшить ручной труд и, следовательно, временные затраты, но не позволил обеспечить приемлемую точность распознавания близких лексем. Например, «Минтай, 1 кг» и «Минтай, ж/б 320 г» попадали в одну категорию «Рыба» при наличии раздела «Консервы».

Для повышения точности классификации было решено использовать сверточную нейронную сеть [1]. Учитывая малый объём входной информации (одно предложение)

использовалась сеть со сверточным, субдискретизирующим и полносвязным слоем.

Для перевода слов в вектор фиксированной длины использовалась технология Word2Vec от Google [2], что позволило не только очистить входную информацию, снизить размерность, получить векторные представления слов определенной длины, но и отказаться от стемминга и(или) лемматизации, а также удаления стоп-слов.

Обучение нейронной сети проведено на следующих данных: число классов 32, размер обучающей выборки (архив данных портала, разнесённых по разделам вручную) 73000, размер тестовой выборки 14600 товаров, для обучения понадобилось 6 эпох. Критерием оценки точности являлся коэффициент, численно равный доле правильно классифицированных объектов.

Для данной задачи коэффициент точности составил 0.83, что значительно выше полнотекстового поиска (0.68). Такая точность классификации для выбранного количества классов приемлема и позволяет сделать вывод об эффективности использования нейронных сетей для классификации текстовой информации.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Kim, Y. Convolutional Neural Networks for Sentence Classification [Electronic resource]. – Mode of access: <https://arxiv.org/pdf/1408.5882.pdf>. – Data of access 29.03.2021 г.
2. Word2Vec: как работать с векторными представлениями слов [Электронный ресурс]. – Режим доступа: <https://neurohive.io/ru/osnovy-data-science/word2vec-vektornye-predstavlenija-slov-dlja-mashinnogo-obucheniya/>. – Дата доступа: 29.03.2021 г.