

Нейросетевой подход к анализу текста

*Мисник Антон Евгеньевич
Игнатова Дарья Сергеевна
Белорусско-Российский университет,
г. Могилев, РБ*

Анализ текстов является достаточно актуальной и в то же время ресурсоёмкой задачей. Существующие подходы в основном ориентированы на сравнительный анализ слов и словосочетаний с эталонными выборками.

Целью данного исследования было проанализировать базовые возможности применения нейронных сетей для кластеризации слов по частям речи, что, как минимум, позволяет ускорить сравнительный анализ.

Для нейросетевого анализа принадлежности слова к той или иной части речи необходимо, в первую очередь, определиться с тем, каким образом преобразовывать слово к виду, пригодному для передачи нейронной сети.

В настоящее время два подхода считаются основными для преобразования слов к пригодному для передачи нейронной сети виду: one hot encoding и word2vec.

One hot encoding предполагает сопоставление каждому элементу выборки, например слову, значения в матрице, размерностью в выборку. Данный подход не очень удобен в ситуации, когда выборка у нас равняется всему текущему словарю

Word2vec – более сложный вариант, уже основывающийся на машинном обучении. Классически, он принимает текст в качестве входных данных и сопоставляет каждому слову вектор, определяя координаты слов в тексте. Данный подход весьма затратен по ресурсам и времени, кроме того, получаемые вектора для конкретного слова, что очевидно, меняются, в зависимости от переданного для анализа текста.

Предложен модифицированный вариант one hot encoding, при котором слово по семантическим алгоритмам разбивается на части, после чего, каждая часть обрабатывается по правилам one hot encoding. При данном подходе размер матриц для каждой части слова существенно уменьшается, скорость обработки возрастает.

При использовании многослойного персептрона в качестве тестовой архитектуры нейронной сети и модифицированного варианта one hot encoding, при анализе слов английского языка на принадлежность к частям речи точность достигла 87%.

К вопросу о роевом интеллекте

*Истомин Валерий Валерьевич
Московский государственный технический
университет им. Н.Э. Баумана (МГТУ)*

Экстремальная робототехника занимается вопросами выполнения задач, поставленных робототехнической системе в экстремальных внеш-