

ИСПОЛЬЗОВАНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ОПРЕДЕЛЕНИЯ ОЦЕНОЧНОЙ КОННОТАЦИИ ТЕКСТА

Потемкин С.Б.

МГУ им. Ломоносова, Россия, 119481 Москва, Ленинский проспект, 99, +79031550543,
prolexprim@gmail.com

При разработке программы анализа оценочного значения необходимо обучить программу на выборке текстов конкретной тематики. Набор лексических единиц, входящих в отзыв, имеющих эмоционально- оценочную коннотацию, можно легко выявить с использованием соответствующих словарей, проведя предварительную лемматизацию лексики. Эти данные, закодированные очевидным способом, будут использованы как входной набор для нейронной сети, а общая оценка, содержащаяся в отзыве, будет результатом работы нейронной сети (НС). Ожидается, что НС определит положительную или отрицательную характеристику всего отзыва. Решение задачи достигается путем ввода последовательности положительно (+1), отрицательно (-1) окрашенных лексем в отзыве в нейронную сеть, с последующей настройкой НС, прогоном НС на этих данных, получения результатов и их интерпретации. Уровень знаний, необходимый для успешного применения нейронной сети, намного меньше, чем, к примеру, при использовании методов регрессионного анализа, экспертных систем, машины SVM, разделяющей гиперплоскости и пр. Результат сравнивается с реальными оценками отзыва. Это сравнение и является материалом для обучения НС. Нейронная сеть с прямым распространением строится вызовом функции Matlab®. $net = newff(P, T, N)$; где P – данные оценочной лексики и T – целевые значения, данные об оценке всего отзыва, N – число нейронов в слое. Входные данные автоматически делятся объектом net Matlab® на обучающий, проверочный и тестовый наборы. Обучение продолжается до тех пор, пока сеть продолжает улучшать соответствие между результатами, полученными на выходе сети и реальными результатами оценки отзыва. Классификация на сырых данных, полученных экспериментально проведена за 6 эпох, валидация и тестирование дали приемлемые результаты: Процент правильной классификации: 72,6%, процент неверной классификации: 28,4% отзывами, Сеть имеет 1 слой, 20 нейронов, функция активации logsig (логистическая функция).

Результатом исследования стало: Нейронная сеть может применяться для практических задач классификации положительных и отрицательных отзывов в зависимости от заданной оценочной лексики. На тестовом примере НС показывает правильные результаты. Результаты не зависят от вида функции активации с насыщением и неприемлемы для линейной функции активации. Необходимо проводить дальнейшие исследования с новыми наборами данных, включающих несколько сотен или даже тысяч отзывов. Интересно получить результаты при градуированных величинах оценочной функции лексемы. Это предполагается сделать в рамках дальнейших исследований.