

Д. И. Гузино, Е. В. Косарева
(УО «ГрГУ им. Я. Купалы», Гродно)

МОДЕЛЬ ИДЕНТИФИКАЦИИ E-MAIL СПАМА НА ОСНОВЕ НЕЙРОННЫХ СЕТЕЙ ПРЯМОГО РАСПРОСТРАНЕНИЯ

В последнее время актуальной является задача идентификации e-mail спама. В данной работе описано применение модели машинного обучения основанной на использовании многослойного персептрона [1]. Алгоритм обучения нейронной сети реализован на языке программирования Python с использованием библиотеки машинного обучения scikit-learn [2].

Данные для обучения и тестирования модели представлены в текстовом документе: первый элемент строки соответствует классу объекта, а остальные элементы – уникальным индексам слов и частоте их появления в электронном письме. Выбор наилучшей модели осуществлялся по следующим параметрам: функция активации нейронной сети, количество скрытых слоев и параметр регуляризации α , который варьируется для того, чтобы предотвратить переобучение. Для нахождения параметров нейронной сети использовался оптимизатор Adam, основанный на методе стохастического градиентного спуска.

Наибольшая точность классификации 0.9943 была получена для нейронной сети с выпрямленной линейной единичной функцией активации при параметрах $\alpha = 10^{-5}$ и количестве нейронов в каждом слое 138, 16 и 2 соответственно.

Литература

1 Хайкин, С. Нейронные сети. Полный курс. Пер. с англ. / Н. Н. Куссуль, А. Ю. Шелестова. 2-е изд., испр. – М.: Издательский дом Вильямс, 2008. – 1103 с.

2 Neural network models [Электронный ресурс]. – 2017. – Режим доступа: http://scikit-learn.org/stable/modules/neural_networks_supervised.html/. – Дата доступа: 12.01.2018.