

ВЫБОР ПРИЗНАКОВОГО ПРОСТРАНСТВА ДЛЯ КЛАССИФИКАЦИИ IP-ТРАФИКА СЕТИ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

Улугбек Рахимжон угли Охундадаев
Национальный университет Узбекистана
ulugbek_1122@mail.ru

АННОТАЦИЯ

IP-протокол и протоколы транспортного уровня (TCP, UDP) имеют множество различных параметров и характеристик, которые можно получить из как непосредственно заголовков пакетов, так и статистических наблюдений за потоками. Для решения задачи классификации сетевого трафика методами машинного обучения необходимо определить набор данных (признаков), которые целесообразно использовать для решения задачи классификации. Выбор признаков зависит от требований к процессу классификации – скорости и точности классификации. В работах зарубежных авторов показано, что можно выделить до 248 различных атрибутов IP-трафика сети, которые потенциально можно использовать в методах машинного обучения для классификации или кластеризации IP-трафика по приложениям. Анализ показывает, что не все предложенные атрибуты одинаково влияют на точность и скорость классификации.

Ключевые слова: классификация, машинное обучение, объект, характеристика, нейронная сеть, управление данными, атрибут, SVM, IP-трафик, протокол.