

СРАВНЕНИЕ И ВЫБОР ОПТИМАЛЬНОГО НЕЙРОСЕТЕВОГО АЛГОРИТМА ФИЛЬТРАЦИИ СПАМ-ПИСЕМ

Процесс фильтрации спам-писем является неотъемлемой частью информационной безопасности сети Интернет. Он достаточно нетривиален и ресурсоемок. Для оптимизации процесса защиты информации используются статистические методы и нейронные сети, которые становятся частью существующей системы защиты. Это одно из перспективных направлений в области защиты информации.

ВВЕДЕНИЕ

Одним из способов решения данной проблемы, проблемы огромного количества спам-писем, является применение алгоритмов фильтрации. На данный момент существует множество механизмов и алгоритмов фильтрации писем, имеющих различную эффективность и имеющие некоторую ненулевую вероятность ложного срабатывания.

Цель исследования состоит в нахождении способа уменьшения времени на обработку спам-писем без потери производительности и уменьшения пропускной способности почтовых систем.

I. СРАВНЕНИЕ АЛГОРИТМОВ

Как уже было упомянуто, существует множество алгоритмов, которых можно разделить на классические алгоритмы и алгоритмы на основе нейронных сетей.

Классические алгоритмы основаны на статистике и теории вероятностей, а алгоритмы на основе нейронных сетей, в свою очередь, используют идеи и методы классических алгоритмов вместе со знаниями, полученными из иных дисциплин.

Среди представителей классических алгоритмов есть наивный Байесовский классификатор. Недостатком классификатора является необходимость иметь вероятности всех возможных слов/фраз из спам-писем. Достоинством является линейное время выполнения.

К алгоритмам на основе нейронных сетей относится многослойный перцептрон. К недостаткам относится расхождение модели при обучении на скачкообразных данных при недостатке слоев или нейронов в этих слоях, что требует для обучения большой сети использования специализированного оборудования. К достоинствам относится способность нейронов находить скрытые

связи в данных, а также и способность обучаться на новых данных без переобучения всей модели. Разновидностью перцептрона является рекуррентная нейронная сеть, позволяющая обрабатывать последовательности данных и событий во времени.

II. СОВМЕЩЕНИЕ АЛГОРИТМОВ

Для достижения оптимальной фильтрации данных необходимо использовать алгоритмы на основе нейронных сетей совместно с классическими алгоритмами.

Идея для оптимального алгоритма фильтрации состоит в последовательном применении классических алгоритмов фильтрации до достижения определённой ошибки нейросетевой модели. После этого, нейронные сети будут использоваться единолично. Данный алгоритм частично похож на ансамбль методов обучения моделей, но использует подобную идею.

ЗАКЛЮЧЕНИЕ

Среди алгоритмов, применяемых для фильтрации данных, а в частности мгновенных сообщений и электронной почты, имеется множество как производительных, но имеющих высокую вероятность ложного срабатывания, так и точных, но медленных алгоритмов.

Выявлено, что нейросетевые алгоритмы имеют высокие требования к аппаратному обеспечению, что возможно компенсировать их совместным использованием с классическими алгоритмами.

1. Rosenblatt, F. F. Principles of Neurodynamic: Perceptrons and the Theory of Brain Mechanisms / F. F. Rosenblatt – 1965. – 480 с.
2. Spam detection using neural networks [Электронный ресурс] / Портал GitHub. – Режим доступа: <https://medium.com/emergent-future/spam-detection-using-neural-networks-in-python-9b2b2a062272>. – Дата доступа: 12.04.2019

Заграй Владимир Юрьевич, магистрант кафедры информационных технологий автоматизированных систем БГУИР, v_id@mail.ru.

Научный руководитель: Гуринович Алеветина Борисовна, заместитель декана ФИТУ по научно-методической работе, кандидат технических наук, доцент, gurinovich@bsuir.by.