

ИССЛЕДОВАНИЕ АЛГОРИТМОВ ФИЛЬТРАЦИИ СПАМ-ПИСЕМ

В настоящее время, век информационных технологий люди используют электронные средства для обмена сообщениями и конфиденциальной информацией, но этому всячески пытаются воспрепятствовать злоумышленники с целью кражи персональной информации.

ВВЕДЕНИЕ

Тема обмена сообщениями, как упоминалось ранее, сейчас очень актуальна. Также упоминалась проблема похищения информации. Существует множество механизмов фильтрации писем, но все они имеют разную эффективность и имеют вероятность ложного срабатывания.

Цель исследования состоит в нахождении способа уменьшения вероятности пропуска спам-писем, отправляемых пользователям посредством сети интернет. Применение результатов исследования позволит уменьшить вероятность пропуска спам-писем. Нахождение такого способа возможно путём исследования множества алгоритмов фильтрации и сравнения их эффективности, времени на калибровку.

I. ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ

Для достижения оптимальной фильтрации данных необходимо использовать алгоритмы на основе нейронных сетей совместно с классическими алгоритмами. Создать оптимальный алгоритм можно с применением нейронных сетей и машинного обучения.

Нейронная сеть (искусственная нейронная сеть) – это система соединенных и взаимодействующих между собой нейронов (простых процессоров).

На текущий момент нейронные сети применяются для распознавания объектов, в том числе и спам-писем.

Для фильтрации спам-писем применяются алгоритмы:

1. наивный байесовский классификатор;
2. обучаемые нейронные сети;
3. метод опорных векторов;
4. генетические алгоритмы.

II. УМЕНЬШЕНИЕ ВЕРОЯТНОСТИ ПРОПУСКА СПАМ-ПИСЕМ

Идея для оптимального алгоритма фильтрации состоит в последовательном применении классических алгоритмов фильтрации, которые

будут удалять письма с некоторым заранее заданным текстом, и последующим использованием нейронных сетей и генетического алгоритма, которые будут удалять сообщения, пропущенные первым слоем с заранее заданными правилами. Также нейронная сеть имеет свойство самообучаться на входящих сообщениях, что позволит удалять часто приходящие сообщения подозрительного содержания.

Схема обработки сообщений с применением нейронной сети приведена на рисунке 1.

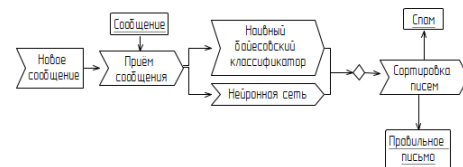


Рис. 1 – Схема обработки сообщений посредством двух алгоритмов

Основной недостаток в использование нейронных сетей заключается в относительно большом времени обучения системы. Но данный недостаток компенсируется использованием классических алгоритмов: даже если нейронная сеть не будет нормально настроена, будут использоваться классические алгоритмы. Когда нейронная сеть будет обучена, она будет использоваться без классических алгоритмов, что позволит ускорить процесс фильтрации и уменьшить вероятность ложных срабатываний.

III. ВЫВОДЫ

Таким образом, комбинация нейронных сетей с классическими алгоритмами позволяет уменьшить количество спам-писем и вероятность их появления, а также увеличивает надежность систем фильтрации в целом.

1. 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: 02.04.2018.

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

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