ФИЛЬТРАЦИЯ СПАМ-ПИСЕМ С ПОМОЩЬЮ АЛГОРИТМОВ НА ОСНОВЕ НЕЙРОННЫХ СЕТЕЙ

Заграй В. Ю., Гуринович А. Б.
Кафедра информационных технологий автоматизированных систем, кафедра вычислительных методов и программирования, Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: v_vld@mail.ru, gurinovich@bsuir.by

Процесс отсеивания спама является неотъемлемой частью информационной безопасности. Он достаточно ресурсоемок и нетривиален. Для оптимизации процесса защиты информации используются нейронные сети, которые становятся частью существующей системы защиты. Это одно из самых перспективных направлений в области защиты данных.

Введение

Одной из главной задач при осуществлении обмена сообщениями и письмами является фильтрация или отсеивание вредноносных сообщений и так сообщений, не содержащих информации, не требующей внимания пользователя, как реклама ненужных товаров или предложения для нелегального заработка, называемые спам-письмами. Алгоритмы фильтрации спам-писем предназначены для отсеивания таких сообщений, но имеют вероятность ложного срабатывания.

Цель работы состоит в нахождении способа уменьшения вероятности пропуска спам-писем, отправляемых пользователям посредством сети интернет. Последующее применение результатов работы позволит уменьшить указанную ранее вероятность.

I. Подготовка данных

Для использования алгоритмов фильтрации для определения спам-писем, необходимо определить ключевые характеристики писем, по которым производится фильтрация.

Такими характеристиками являются:

- Повторяющиеся слова;
- словосочетания;
- общие количества символов, знаков пунктуации, цифр, пробельных символов, слов;
- частота каждой буквы, специальных символов;
- средние длины слова, предложения;
- распределение частоты длин слов;
- мера разнообразия;
- количество уникальных слов.

Для получения характеристик используется словарь, содержащий в себе слова, символы и прочие элементы текста, наиболее характерные для спама.

Выбор ключевых характеристик напрямую влияет на выбор конкретного алгоритма фильтрации. У некоторых алгоритмов конечный результат зависит от начальных данных, т.е. при неправильном выборе характеристик писем можно увеличить вероятность ложного срабатывания. Исходя из этих характеристик, можно оценивать текст на принадлежность к спаму комплексно, полагаясь на множество разнородных параметров, которые дополняют друг друга и уточняют оценку при принятии решения.

II. Выбор алгоритмов

Для выполнения фильтрации спам-писем необходимо полученные ранее характеристики сообщений обработать алгоритмами классификации текста и данных. Существуют классические алгоритмы и алгоритмы на основе искусственных нейронных сетей.

Классические алгоритмы представляют собой методы и алгоритмы построеные на применении методов статистического анализа данных и математических вычислений. К ним относятся:

- Наивный байесовский классификатор;
- метод k-ближайших соседей;
- метод опорных векторов;
- генетические алгоритмы.

Алгоритмы фильтрации на основе искусственных нейронных сетей предназначены на разделения задачи на блоки. При этом каждый блок обрабатывается более простыми процедурами, чем в классических алгоритмах, что позволяет динамически расширять сети без нарушения последовательности выполнения алгоритма. К эти алгоритмам относятся:

- Распознавание образов;
- перцептрон;
- нейронная сеть Кохонена;
- самоорганизующаяся карта Кохонена.

Применение алгоритмов на основе необученных нейронных сетей приводит к нежелаемым результатам, т.е. к пропуску спам-писем. Данный недостаток возможно минимизировать путём совместного использования классических алгоритмов и алгоритмов на основе нейронных сетей на начально этапе обучения или самообучения сети.

Как было указано в предыдущем разделе, из исходных сообщений, вероятно, содержащих спам-письма, извлекаются отдельные данные, представляющие собой количественные характеристики сообщений. Данные характеристи-

ки возможно обрабатывать параллельно с целью уменьшения времени выполнения фильтрации.

III. Обработка ключевых характеристик

При фильтрации спам-писем примененяют классические алгоритмы, так как они уже достаточно хорошо изучены. Одним из них является Наивный байесовский классификатор. Его суть состоит в определении вероятности того, что конкретное письмо является спамом. Одним из недостатков является тот факт, что необходимо заранее определить набор ключевых характеристик, описывающих спам-письма, которого лишены самообучаемые нейронные сети. Преимуществом является быстрое время выполнения. При обучении алгоритма вычисляются вероятности появления отдельных характеристик в спам-письмах формула: $P(C) = \frac{m}{n}$, где m - это количество спам-писем, n - количество всех писем с характеристикой C

Также применяются неклассические алгоритмы. Анализ ключевых характеристик нейронной сетью напоминает байесовскую фильтрацию спама, где для каждого слова или словосочетания можно установить коэффициент определения письма как спам. Однако, в отличие от байесовского фильтра, здесь коэффициенты - это веса между нейронами сети, способные динамически изменяться в процессе обучения, что позволяет эффективно обнаруживать новый и ранее неизвестный спам за счет умения нейронной сети обобщать накопленный опыт. Нейронные сети внешне похожи на Наивный байесовский классификатор, но структурно различаются.

Такую нейронную сеть можно структурно реализовать в виде многослойного перцептрона со скрытыми слоями или в виде сети Кохонена.

Перцептронная сеть проста в реализации и представляет собой перцептрон с числом входных параметров n, равных размерности входного вектора характеристик (при использовании полученных выше характеристик $n = 7$). Она имеет единственный нейрон, выдающий значение вероятности обнаружения спама в тексте, принимающий значение от 0 до 1. Перцептронная сеть выполняет единственную функцию – принятие решения о наличии спама.

Также может быть применена сеть Кохонена, которая выполняет кластеризацию данных, что позволяет эффективнее определить направленность текста, в том числе отбросить на этапе кластеризации текст, являющийся обычным текстом. Нейрона сети Кохонена в базовом варианте имеет вид:

$$y_j = w_{j0} + \sum_{i=1}^{m} w_{ji} x_i$$

В данной формуле w_{ji} - это вес i-й характеристики j-ого нейрона.

Обычно применяется немаленькое число нейронов для нахождения "лучшего"нейрона, имеющего наибольший выход. Полученное на выходе значение сверяется с пороговой величиной, определяющей границу между спамом и обычной почтой.

Нейронные сети, как и биологические организмы в начале своего существования, являются необученными, то есть не могут использоваться без обучения. Для минимизации ошибок от использования недостаточно обученных нейронных сетей следует применять классические алгоритмы, так как они показывают меньшую вероятность ложного срабатывания по сравнению с нейронной сетью. Однако впоследствие нейронные сети, обученные на реальных данных и результатах, полученных от классических алгоритмов, имеют меньшую вероятность ложного срабатывания при высокой производительности, чем классические алгоритмы.

Заключение

Среди алгоритмов, применяемых для фильтрации данных, а в частности электронной почты и сообщений, имеется множество как производительных, но имеющих высокую вероятность ложного срабатывания, так и точных, но медленных алгоритмы. Совместное использование нейронных сетей с классическими алгоритмами позволяет уменьшить количество спам-писем, а также уменьшить вероятность их пропуска фильтром.

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: 20.09.2018
3. Statistical Comparisons of Classifiers over Multiple Data Sets [Электронный ресурс] -– Режим доступа: http://sci2s.ugr.es/sicidm/pdf/2006-Demsar-JMLR.pdf. — Дата доступа: 23.09.2018
4. Нейронные сети: полный курс / С. В. Хайкин [и др.]. – М: Вильямс, 2006. – 1104 с.
5. Ясницкий, Л. Н. Введение в искусственный интеллект / Л. Н. Ясницкий –М: Издательский центр «Академия», 2010. – С. 176.
6. Андреев, А. М. Автоматическая классификация текстовых документов с использованием нейросетевых алгоритмов и семантического анализа. Режим доступа: http://www.inteltec.ru (дата обращения: 10.09.2018).
7. Garshin, A. A. An automated system of recognizing handwritten digits based on convolution neural networks / A. A. Garshin, O. P. Soldatova – Register of computer programs – 2010
8. Головко, В. А. Нейронные сети: обучение, организация и применение / В. А. Головко // М.: ИПРЖР, 2001. – 205 с.