Интернет-журнал «Науковедение» ISSN 2223-5167 http://naukovedenie.ru/

Том 8, №3 (2016) http://naukovedenie.ru/index.php?p=vol8-3

URL статьи: http://naukovedenie.ru/PDF/04TVN316.pdf

Статья опубликована 23.05.2016.

Ссылка для цитирования этой статьи:
Ларионова А.В., Хорев П.Б. Метод фильтрации спама на основе искусственной нейронной сети // Интернет-журнал «НАУКОВЕДЕНИЕ» Том 8, №3 (2016) http://naukovedenie.ru/PDF/04TVN316.pdf (доступ свободный). Загл. с экрана. Яз. рус., англ.

УДК 004.81

Ларионова Анна Владимировна[1]
ФГБОУ ВО «Российский государственный социальный университет», Россия, Москва[2]
Аспирант
E-mail: tarelo4ka76@mail.ru
РИНЦ: http://elibrary.ru/author_items.asp?authorid=828879

Хорев Павел Борисович
ФГБОУ ВО «Национальный исследовательский университет «Московский энергетический институт», Россия, Москва
Преподаватель
Кандидат технических наук, доцент
E-mail: pbkh@mail.ru
РИНЦ: http://elibrary.ru/author_items.asp?authorid=620811

Метод фильтрации спама на основе искусственной нейронной сети

Аннотация. В данной статье рассматривается задача фильтрации спама и наиболее распространенные подходы к ее решению: на основе списков адресов, сигнатур, теоремы Байеса в сравнении с методами искусственного интеллекта. Задача фильтрации спама является актуальной проблемой, так как технологии создания спама развиваются следом за средствами защиты от спама, что требует переосмысления подходов к задаче фильтрации спама и применения методов и средств искусственного интеллекта. В качестве решения проблемы предлагается использовать подход на основе методов искусственного интеллекта, в частности на основе искусственной нейронной сети. Данный подход требует подготовки обучающей и тестовой выборок сообщений для обучения классификатора, выделения значимых признаков сообщений, настройки параметров модели, оценки точности классификатора.

Ключевые слова: фильтрация спама; искусственные нейронные сети; искусственный интеллект; спам; классификация сообщений; выделение признаков сообщений; персептрон; теорема Байеса; машинное обучение; информационная безопасность

В современном мире, где реклама является двигателем торговли, с развитием сети Internet и средств общения, проблема нежелательной рекламы [6] и сообщений требует интеллектуального подхода для ее решения. Современные методы борьбы со спамом, основанные на лингвистических сигнатурах, правилах фильтрации сообщений, становятся все

[1] LinkedIn: https://ru.linkedin.com/in/anna-larionova-74434689
[2] 140180, Российская Федерация, Московская область, г. Жуковский, ул. Семашко, д. 8, корп. 2, кв. 41

менее эффективными [6, 8], так как требуется увеличение трудозатрат специалистов по защите от спама на поддержание этих сигнатур и правил в актуальном состоянии. Таким образом, современные методы борьбы со спамом требуют постоянного участия человека для эффективного анализа текста, они не способны самостоятельно вырабатывать эти правила, то есть самообучаться. Если рассматривать человека как средство борьбы со спамом, то можно сказать, что он обладает способностью обнаружения признаков спама, основываясь на собственном опыте и предпочтениях, знаниях о добровольных новостных и рекламных подписках, обучаемостью, его работа не сводится к шаблонам и потому более эффективна. Именно поэтому задача создания средства борьбы со спамом сводится к наделению средства борьбы со спамом навыками и качествами, присущими человеку: способность к обучению, система предпочтений и исключений, анализ контекста, система принятия решений.

Предлагаемый автором статьи метод фильтрации спама основан на использовании нейронной сети, выступающей в качестве механизма принятия решений, давая на выходе вероятностную оценку «спамности» всего сообщения. Искусственная нейронная сеть обладает способностью обучаться (в том числе, обобщать свои знания, накапливать опыт), является наиболее приближенной моделью человеческого мозга, как по архитектуре, так и по принципам работы. Более подробно читатель может ознакомиться с свойствами искусственной нейронной сети и моделированию человеческой деятельности и мышления в [1, 2, 3, 4, 7, 9, 10].

Задача обеспечения информационной безопасности, куда входит и защита от спама, является нетривиальным ресурсоемким процессом. Для увеличения эффективности и повышения степени автоматизированности процесса защиты информации с целью освобождения человеческих ресурсов наиболее перспективным направлением является внедрение нейросетевых технологий в существующие системы защиты. Так, например, нейронные сети получили широкое распространение в системах обнаружения и отражения сетевых атак. В таких системах, как и в предлагаемом автором статьи методе, нейронные сети анализируют комплекс разнородных параметров сети (время ответа сервера, отклонение пакетов от стандартов RFC и прочее), выявляя аномальное поведение и способны опознавать даже те атак, которых не было в обучающей выборке благодаря способности нейронных сетей к обобщению и обучению. Предлагаемая автором статьи система фильтрации спама принципиальна схожа с подобными системами обнаружения и отражения атак, только обнаруживаются не сетевые атаки, а сообщения, чье содержание является нежелательной рекламой (спамом). Отличие состоит в том, что предлагаемая система фильтрации спама работает на уровне приложения, согласно модели OSI, а не на уровне сети, как системы обнаружения атак, и анализируются не флаги пакетов, а непосредственно данные, т.е. содержание сообщений.

Любая нейронная сеть имеет входы, выходы, собственно нейроны и связи между ними (синапсы, аксоны, дендриты) [4]. Схема простой нейронной сети изображена на рисунке 1.

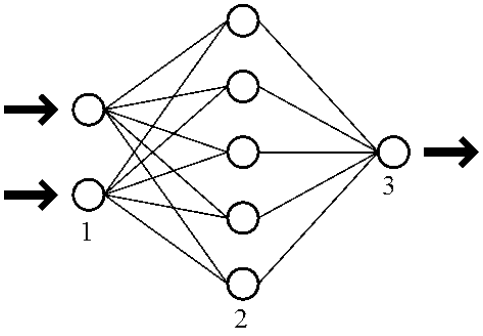

Рисунок 1. *Схема нейронной сети. 1 – нейроны входного слоя; 2 – нейроны скрытого слоя; 3 – нейрон выходного слоя*

На примере человека рассмотрим процесс обнаружения спама в сообщении. Во-первых, существует ряд слов и словосочетаний, которые довольно часто встречаются в спаме (например, «buy *something* with 50% discount»). Однако, это не является достаточным основанием для отнесения такого сообщения к спаму. В этом случае человек дополнительно обращает внимание на контекст и смысл сообщения, его общую направленность, также он может обратить внимание на орфографические, синтаксические и морфологические особенности текста. Исходя из этой совокупности, с гораздо большей достоверностью можно принять решение о том, является ли данное сообщение спамом или нет. Поэтому входными параметрами нейронной сети в составе средства борьбы со спамом являются предварительно выявленные статистические признаки сообщения:

- удельное число слов с подозрением на спам в сообщении;
- удельное число словосочетаний и фраз с подозрением на спам в сообщении.

А также нестатистические входные параметры:

- семантические признаки;
- направленность текса;
- морфологические признаки – правильность построения предложения и установления связей между частями речи (формализм Бекуса-Наура) [4];
- орфографические признаки – правильность написания слова, наличие замены сходных по написанию символов (например, замена буквы «О» цифрой «0» для обмана сигнатурной (шаблонной) фильтрации).

Поскольку нейронная сеть оперирует численными значениями, необходимо сформировать из вышеописанных признаков числовой входной вектор значений [2].

Для получения статистических признаков используется специальный словарь, содержащий в себе слова, наиболее характерные для спама. В исходном сообщении производится поиск и подсчет слов, которые совпадают с содержимым данного словаря. Для улучшения точности принятия решений дополнительно производится подсчет наиболее часто употребляемых в спаме словосочетаний. Это уменьшает вероятность ложного срабатывания.

Анализ статистических признаков нейронной сетью напоминает байесовскую фильтрацию спама [8, 9], где для каждого слова или словосочетания можно установить коэффициент «спамности». Однако, в отличие от байесова фильтра, здесь коэффициенты - синаптические связи (веса) между нейронами, способные динамически изменяться в процессе обучения, что позволяет эффективно обнаруживать новый и ранее неизвестный спам за счет умения нейронной сети обобщать накопленный опыт. Таким образом, внешне нейронная сеть будет схожа с байесовым фильтром, однако, они различаются внутренней архитектурой, дополнительными функциями и свойствами нейронной сети: нейронная сеть не зависит от формы представления данных и способна обрабатывать семантические, фонетические и орфографические признаки, если представить их в виде числовых значений. Исходя из этого, можно оценивать текст на принадлежность к спаму комплексно, полагаясь на множество разнородных параметров, которые дополняют друг друга и уточняют оценку при принятии решения.

Данную нейронную сеть можно структурно реализовать в виде многослойного персептрона [7, 10] со скрытыми слоями или используя гибридную нейронную сеть на основе

сети Кохонена [3] и персептрона [7]. Первый случай наиболее прост в реализации и представляет собой персептрон с числом входных параметров n, равных размерности входного вектора (в нашем случае n=5 или n=6, если учитывать словосочетания). Он будет иметь единственный выходной нейрон, выдающий значение вероятности обнаружения спама в тексте, принимающий значение от 0 до 1. Данная нейронная сеть будет выполнять единственную функцию – принятие решения о наличии спама. Во втором случае сеть Кохонена выполняет кластеризацию входных параметров [2], что позволит эффективнее определить направленность текста, в том числе отсеять текст, не являющийся спамом на этапе кластеризации. Карта Кохонена способна к обучению без учителя, что уменьшает временные затраты на обучение нейронной сети. Роль персептрона в данной гибридной нейронной сети также сводится к процессу принятия решения о наличии спама в сообщении. В первом случае в качестве активационной функции используется одна из сигмоидных функций [4], во втором случае, помимо сигмоидной функции активации, используется функция Гаусса [4].

Нейронная сеть неспособна непосредственно оценивать текст на наличие спама, поскольку оперирует числовыми значениями. Также надо учесть, что сам текст сообщения может содержать орфографические и синтаксические ошибки, которые затрудняют процесс анализа, поэтому их необходимо предварительно обнаружить и исправить, затем выделить из текста входные параметры нейронной сети. Для выделения этих параметров необходимо использовать синтаксический анализ предложений. Синтаксический анализ автор статьи предлагает производить, основываясь на формализме Бекуса-Наура, где в качестве инструментария используется база данных, которая представляет собой словарь с морфологической и орфографической оценками и общей семантикой.

Исходя из этого, можно выделить следующие сопутствующие технологии:

- синтаксический анализатор текста;
- база данных;
- статистический анализ текста.

Как уже было сказано, для увеличения точности метода, необходимо произвести первичную обработку текста, в ходе которой исправляются орфографические ошибки в словах, устраняются лишние пробелы и выделяются слова и предложения из текста, заменяются «обманные» символы (например, цифра «0» заменяется буквой «О»).

Далее производится вторичная обработка текста, на этапе которой осуществляется формализация текста с помощью правил Бекуса-Наура, выделение признаков спама, формирование входного вектора для нейронной сети.

В отличие от байесовой фильтрации спама, предложенный автором статьи метод учитывает наличие в сообщении «обманных» символов, которые обычные фильтры пропускают. В предложенном методе фильтрации спама используется множество разнородных параметров, то есть не только статистические (лексические), но и морфологические и синтаксические, наличие орфографических ошибок в словах и ошибок при построении предложений. Данная система способна к самообучению, обнаружению ранее неизвестных спам-сообщений, в то время как эффективность байесова фильтра зависит от постоянной коррекции коэффициентов на новых выборках [5], нет процесса самообучения. Для каждого нового спам-сообщения при использовании байесова фильтра необходимо корректировать коэффициенты «спамности», а при использовании фильтрации на основе шаблонов необходимо постоянно пополнять базу шаблонов, то есть содержать специалистов, которые будут поддерживать актуальность этой базы. Предложенный автором статьи метод избавлен от многих недостатков байесова фильтра, однако, эффективность метода зависит от

обучающей выборки, используемой в процессе обучения. В итоге возникает задача правильного формирования обучающей выборки, обладающей репрезентативностью и достоверностью. Данная задача вполне выполнима, несмотря на свою сложность, но, будучи однажды выполненной, обеспечивает эффективную работу системы, не требующей постоянного дообучения.

ЛИТЕРАТУРА

1. Осовский Станислав. Нейронные сети для обработки информации = Sieci neuronowe do przetwarzania informacji (польск.) / Перевод И.Д. Рудинского. - М.: Финансы и статистика, 2004. - 344 с. - ISBN 5-279-02567-4.
2. Савельев А.В. На пути к общей теории нейросетей. К вопросу о сложности // Нейрокомпьютеры: разработка, применение. - 2006. - №4-5. - С. 4-14. Режим доступа http://www.radiotec.ru/catalog.php?cat=jr7 (открытый).
3. Хайкин С. Нейронные сети: полный курс = Neural Networks: A Comprehensive Foundation. 2-е изд. - М.: Вильямс, 2006. - 1104 с.
4. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 3-е издание, 2010 – 176 с.
5. Demsar Janez. Statistical Comparisons of Classifiers over Multiple Data Sets - 2006. Access link: http://sci2s.ugr.es/sicidm/pdf/2006-Demsar-JMLR.pdf.
6. Mueller Scott Hazen. What is spam? Information about spam. Abuse.net. Retrieved 2007-01-05. Access link: http://spam.abuse.net/overview/whatisspam.shtml (open access).
7. Rosenblatt, Frank. Principles of Neurodynamic: Perceptrons and the Theory of Brain Mechanisms. - М.: Мир, 1965. - 480 с.
8. Vangelis Metsis, Ion Androutsopoulos, Georgios Paliouras. Spam Filtering with Naive Bayes - Which Naive Bayes? // Third Conference on Email and Anti-Spam (CEAS). 2006 – 9 p.
9. Vapnik Vladimir N. The Nature of Statistical Learning Theory – 1999. Access link http://web.mit.edu/6.962/www/www_spring_2001/emin/slt.pdf (open access).
10. Warren S. McCulloch, Walter H. Pits. A logical calculus of the ideas immanent in nervous activity. Access link http://www.cse.chalmers.se/~coquand/AUTOMATA/mcp.pdf (open access).

Larionova Anna Vladimirovna
Russian State Social University, Russia, Moscow
E-mail: tarelo4ka76@mail.ru

Khorev Pavel Borisovich
National research institute «Moscow Power Engineering Institute», Russia, Moscow
E-mail: pbkh@yandex.ru

Spam filtering method based on artificial neural network

Abstract. This article discusses the problem of spam filtering and the most common approaches to deal with it: based on address lists, signatures, Bayes' theorem in comparison with the methods of artificial intelligence. Spam filtering task is an actual problem, since the technology of spam are advancing together with spam, which requires a rethinking of approaches to the problem of spam filtering and application of artificial intelligence methods. As a solution is proposed to use techniques based on artificial intelligence approach, in particular based on an artificial neural network. This approach requires the preparation of the training and test samples of messages for training the classifier, extraction important features of messages, setting parameters of the model, evaluate the accuracy of the classifier.

Keywords: spam filtering; artificial neural network; artificial intelligence; spam; message classification; feature extraction of messages; perceptron; Bayes theorem; machine learning; information security

REFERENCES

1. Osovskiy Stanislav. Neyronnye seti dlya obrabotki informatsii = Sieci neuronowe do przetwarzania informacji (pol'sk.) / Perevod I.D. Rudinskogo. - M.: Finansy i statistika, 2004. - 344 s. - ISBN 5-279-02567-4.
2. Savel'ev A.V. Na puti k obshchey teorii neyrosetey. K voprosu o slozhnosti // Neyrokomp'yutery: razrabotka, primenenie. - 2006. - №4-5. - S. 4-14. Rezhim dostupa http://www.radiotec.ru/catalog.php?cat=jr7 (otkrytyy).
3. Khaykin S. Neyronnye seti: polnyy kurs = Neural Networks: A Comprehensive Foundation. 2-e izd. - M.: Vil'yams, 2006. - 1104 s.
4. Yasnitskiy L.N. Vvedenie v iskusstvennyy intellekt. M.: Izdatel'skiy tsentr «Akademiya», 3-e izdanie, 2010 – 176 s.
5. Demsar Janez. Statistical Comparisons of Classifiers over Multiple Data Sets - 2006. Access link: http://sci2s.ugr.es/sicidm/pdf/2006-Demsar-JMLR.pdf.
6. Mueller Scott Hazen. What is spam? Information about spam. Abuse.net. Retrieved 2007-01-05. Access link: http://spam.abuse.net/overview/whatisspam.shtml (open access).
7. Rosenblatt, Frank. Principles of Neurodynamic: Perceptrons and the Theory of Brain Mechanisms. - M.: Mir, 1965. - 480 s.
8. Vangelis Metsis, Ion Androutsopoulos, Georgios Paliouras. Spam Filtering with Naive Bayes - Which Naive Bayes? // Third Conference on Email and Anti-Spam (CEAS). 2006 – 9 p.
9. Vapnik Vladimir N. The Nature of Statistical Learning Theory – 1999. Access link http://web.mit.edu/6.962/www/www_spring_2001/emin/slt.pdf (open access).
10. Warren S. McCulloch, Walter H. Pits. A logical calculus of the ideas immanent in nervous activity. Access link http://www.cse.chalmers.se/~coquand/AUTOMATA/mcp.pdf (open access).