

ОБ ИСПОЛЬЗОВАНИИ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ РАСПОЗНАНИЯ СПАМА В ИНТЕРНЕТ

© 2020 А. Г. Юрочкин, Н. А. Коростелева

РАНХиГС при Президенте Российской Федерации (Воронеж, Россия)

*Открытое акционерное общество «Воронежский завод полупроводниковых приборов-сборка»
(Воронеж, Россия)*

В данной работе проводится анализ возможностей методов машинного обучения для борьбы со спамом в глобальной сети Интернет.

Ключевые слова: машинное обучение, спам, интернет.

В настоящее время можно наблюдать развитие различных систем борьбы с распространением спама в интернет. Рассмотрим возможности применения для этих целей алгоритмов машинного обучения [1, 2].

Среди признаков, демонстрирующих появление спама в сети интернет можно отметить следующие:

1. Наблюдение множественных аккаунтов.
2. В сообщениях можно увидеть ссылки, которые связаны с вредным содержанием.
3. Пользователи стремятся ссылки размещать в своих статусах.
4. По ссылкам можно столкнуться с множественными упоминаниям.
5. Существуют жалобы о спаме со стороны большего числа пользователей. Методы борьбы со спамом уже разрабатываются довольно давно. Прежде всего, это было связано с классификации сообществ в электронной почте [3, 4]. Есть недостатки, связанные с SMTP. Также применяется широкий спектр способов машинного обучения.

Перспективными можно считать способы машинного обучения с учителем. При этом могут рассматриваться различные признаки.

Есть такие, которые демонстрируют связи между пользователями и соседями. Есть признаки, имеющие связь с профилями пользователей [5].

Существуют признаки, в которых демонстрируется статус пользователей.

В первых из них осуществляется анализ по числу друзей, подписчиков, соотношений между ними.

Во вторых рассматривается число статусов, дата, когда был создан аккаунт, наличие признаков спама в названиях аккаунтов.

В третьих дается анализ по статусам ответом от пользователей [6, 7].

Среди используемых классификаторов по обнаружению спама весьма популярным считается Random Forest.

Также существуют подходы, позволяющие осуществлять процесс обнаружения спама, который базируется на алгоритме СИА. А в нём проводится сам анализ числа связи между спамерами.

При построении системы, дающей возможности для обнаружения спам-атак, необходимо опираться на назначенные выше подходы, а также проводить анализ списка спам-слов.

Среди методов, позволяющих использовать шаговое обучение в рассмотрении задачи довольно часто прибегают к наивному байтесовскому классификатору, методу опорных векторов и случайному лесу [8, 9].

В первом из этих подходов предоставляется возможности для того, чтобы предсказать класс объектов.

Во втором осуществляется построение оптимальной гиперплоскости в пространстве признаков. В третьем идёт предсказание класса объектов.

Если полученные данные являются несбалансированными, то эффективность применения указанных выше методов заметным образом будет снижена.

Юрочкин Анатолий Геннадьевич – РАНХиГС при Президенте Российской Федерации, доктор техн. наук, профессор, yugoschikinang30@yandex.ru.
Коростелева Наталья Александровна – Открытое акционерное общество «Воронежский завод полупроводниковых приборов-сборка», специалист, kor671steler@yandex.ru.

Таким образом, следует стремиться к повышению эффективности методов обнаружения спама.

На рисунке приведен пример комплексного алгоритма, базирующегося на

машинном обучении, на основе которого можно бороться с проблемами спама в глобальной сети Интернет.

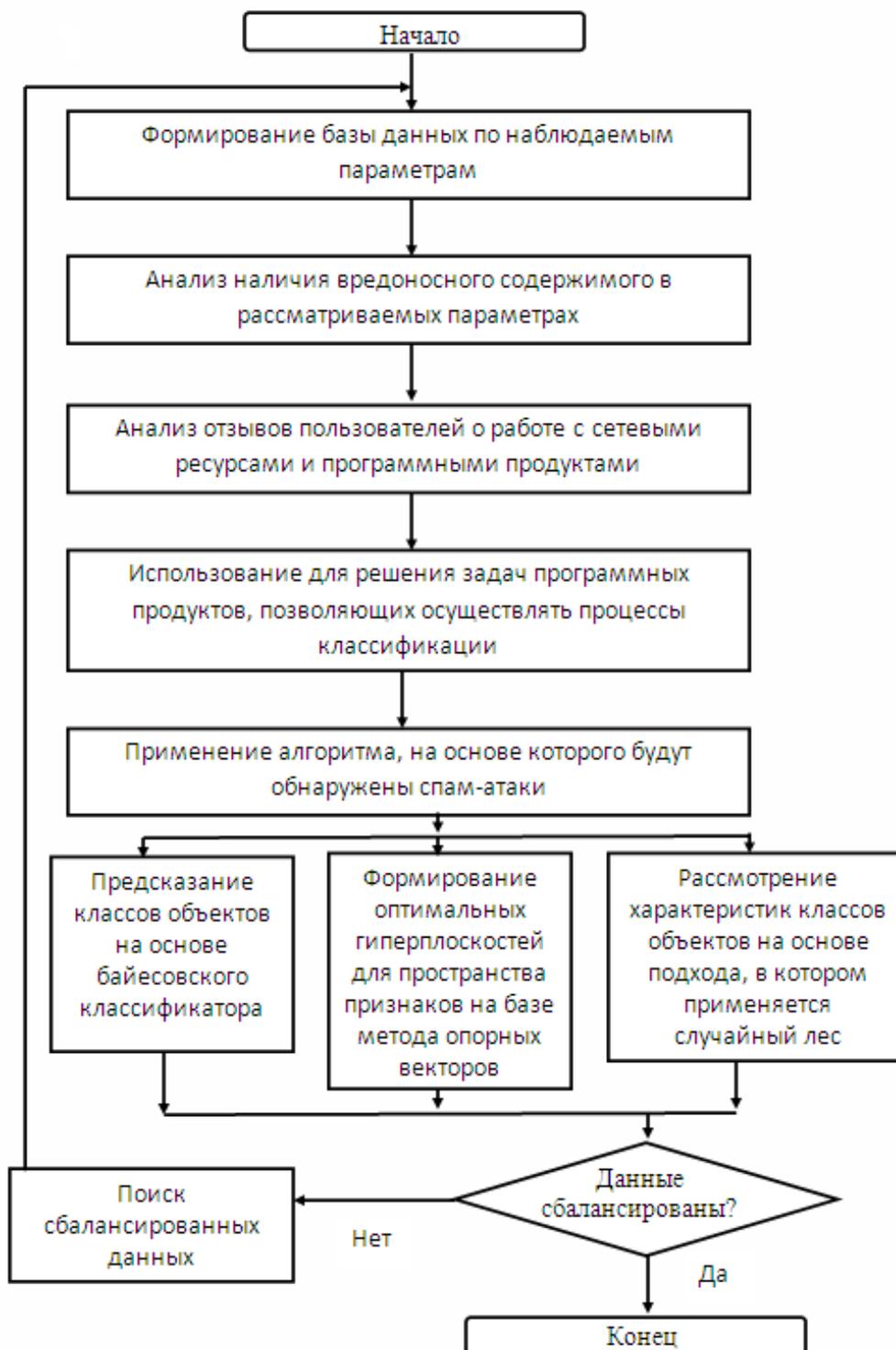


Рисунок. Пример комплексного алгоритма, базирующегося на машинном обучении и для борьбы со спамом.

ЛИТЕРАТУРА

1. Львович, Я. Е. Анализ подходов при проектировании корпоративных информационных систем / Я. Е. Львович // Совре-

менные проблемы экономики и менеджмента. Материалы международной научно-практической конференции. – 2019. – С. 191-193.

2. Преображенский, Ю. П. Информационные технологии в системе интернет-вещей / Ю. П. Преображенский // Интеллектуальные информационные системы. Труды Международной научно-практической конференции. В 2-х частях. – 2019. – С. 25-26.

3. Львович, Я. Е. Проблемы построения корпоративных информационных систем на основе web-сервисов / Я. Е. Львович, И. Я. Львович, Н. В. Волкова // Вестник Воронежского государственного технического университета. – 2011. – Т. 7. – № 6. – С. 8-10.

4. Львович, И. Я. Основы информатики / И. Я. Львович, Ю. П. Преображенский, В. В. Ермолова. – Воронеж, Издательство: Воронежский институт высоких технологий (Воронеж). – 2014. – 339 с.

5. Преображенский, Ю. П. Проблемы цифровизации в современном обществе / Ю. П. Преображенский // Инновационные доминанты социально-трудовой сферы: экономика и управление. Материалы ежегодной международной научно-практической конференции по проблемам социально-трудовых отношений. Редакционная коллегия: А. А. Федченко, О. А. Колесникова. – 2019. – С. 243-245.

6. Гончарова, Н. П. Применение методов аппроксимации данных / Н. П. Гончаро-

ва // Успехи современного естествознания. – 2011. – № 7. – С. 97а.

7. Преображенский, Ю. П. О видах информационных систем в организации / Ю. П. Преображенский // Молодежь и системная модернизация страны. Сборник научных статей 3-й Международной научной конференции студентов и молодых ученых. В 4-х томах. Ответственный редактор А. А. Горохов. – 2018. – С. 131-134.

8. Lvovich, I. Ya. Modeling of information processing in the internet of things at agricultural enterprises / I. Ya. Lvovich, Ya. E. Lvovich, A. P. Preobrazhenskiy, Yu. P. Preobrazhenskiy, O. N. Choporov // IOP Conference Series: Earth and Environmental Science Krasnoyarsk Science and Technology City Hall of the Russian Union of Scientific and Engineering Associations. – 2019. – С. 32029.

9. Львович, И. Я. Исследование алгоритмов распознавания рукописного текста / И. Я. Львович, А. П. Преображенский, О. Н. Чопоров, А. В. Боева, К. В. Сычев // Радиоэлектронные устройства и системы для инфокоммуникационных технологий – РЭУС-2019 Доклады Всероссийской конференции (с международным участием). Сер. «Научные конференции, посвященные дню Радио» 2019. С. 153-156.

ABOUT USE OF MACHINE LEARNING METHODS FOR RECOGNIZING SPAM ON THE INTERNET

© 2020 A. G. Yurochkin, N. A. Korosteleva

RANIGS under the President of the Russian Federation (Voronezh, Russia)

Open joint stock company «Voronezh plant of semiconductor devices-Assembly» (Voronezh, Russia)

This paper analyzes the possibilities of machine learning methods to combat spam on the global Internet.

Key words: machine learning, spam, the Internet.