ПРОБЛЕМЫ АВТОМАТИЧЕСКОЙ ОБРАБОТКИ ТЕКСТОВ

Р. В. Детскина, К. П. Антоник (*Минск, МГЛУ*)

ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ФИЛЬТРАЦИИ СПАМА

В статье рассматриваются вопросы структуры текстов спам-сообщений, применения нейронных сетей для фильтрации спама, этапы нейросетевой технологии фильтрации спама. Анализ статистических признаков нейронной сетью напоминает байесовскую фильтрацию спама. В итоге была создана компьютерная программа, использующая наивный байесовский классификатор, который способен ответить на вопрос, к какой категории классов «спам» – «не спам» относится электронное сообщение.

К л ю ч е в ы е с л о в а: спам-сообщение; фильтрация спама; нейросетевая технология фильтрации спама; автоматизация; компьютерная программа.

The article deals with the structure of spam messages, the use of neural networks for filtering spam, the stages of neural network spam filtering technology. A neural network analysis of statistical features resembles Bayesian spam filtering. As a result, a computer program has been created that uses a naive Bayesian classifier that is able to answer the question, to which category of «spam» – «not spam» classes belongs an email message.

K e y w o r d s: spam message; spam filtering; neural network spam filtering technology; automation; computer program.

Существуют различные методы борьбы со спамом, но ни один из них не дает стопроцентной гарантии защиты от нежелательных рассылок. Именно лингвистический анализ спама имеет значительную практическую ценность. Для разработки эффективных программных средств фильтрации необходимы адекватные данные о структуре и содержании текстов, а также о коммуникативных особенностях отправителей и получателей сообщений. Для создания модели фильтрации спама могут использоваться как более простые алгоритмы, основанные на анализе содержащихся в тексте слов, так и обучаемая нейронная сеть, однако для создания такой модели нужно выделить различные лексические и синтаксические особенности спам-сообщений. Все вышесказанное говорит об актуальности данного исследования.

Одной из наиболее характерных особенностей современного информационного общества, как глобальной системы, является лавинообразно нарастающий процесс его виртуализации. В эпоху электронных коммуникаций рождаются такие важные характеристики коммуникативной культуры, как анонимность коммуникаторов, горизонтальная организованность коммуникации, высокий уровень обратной связи, высокая скорость доступа к информации. Стремительное развитие компьютерных сетей расширило границы доступной информации. Развитие технологий влияет на мировоззренческие позиции людей. Интернет, являясь совершенно новым

пространством коммуникации, создает технологическую основу для формирования культурных сообществ самого разного типа от континентальных и национальных до региональных. В контексте формирования и трансформации культурно коммуникационных систем Интернет становится социокультурным феноменом, активно влияющим на содержательное наполнение, специфику способов осуществления коммуникации и культурно познавательные процессы, в связи с чем интернет-коммуникации выступают эффективным средством освоения культурного наследия. По мнению некоторых исследователей, «…интерпретация практик культурной идентификации в социальных сетях предстает как важная задача социального познания: новые идентичности навязывают свой вариант социального мира, выход в понимание глобальных и международных отношений в мире, обозначают позиции в нем российского сообщества. Культурная идентификация – это практика созидания, воплощения и символического размещения различий, мобилизующих коллективные идентичности» [3].

Интернет, электронные библиотеки, базы данных наряду с традиционной библиотекой представляют собой новые способы хранения, обработки и распространения информации. Если напечатанный текст существовал, если его читали и обсуждали, то экранная культура предполагает только циркуляцию информации. Здесь происходит автоматическая деконструкция текста, его целостность и архаичность уступает место бесчисленному многократному использованию отдельных смысловых элементов. Читатель в Интернете выступает партнером по диалогу. Пространственная и временная характеристика диалога утрачивают свою значимость, ибо информация появляется практически мгновенно и ее распространение больше не ограничивается пространственными барьерами.

Под виртуальной коммуникацией подразумевается компьютерно-опосредованное общение, участником которого может стать каждый пользователь Интернета. Виртуальная коммуникация может осуществляться посредством электронной почты, чатов, форумов.

Спам – это коммуникативная стратегия, которая ориентирована на манипулирование получателем сообщения. Она связана с предоставлением нерелевантной для него информации и навязыванием ему идей манипулятора [5]. Иногда даже человеку сложно определить, является ли сообщение спамом. Совместное использование множества способов фильтрации корреспонденции существенно увеличивает их эффективность. Однако для любого алгоритма фильтрации существует вероятность удаления вместе с нежелательными сообщениями также некоторого количества сообщений, содержащих значимую для получателя информацию. Кроме того, спам многоязычен, и корректная сортировка особенно важна для бизнес-пользователей, ведущих обширную переписку с зарубежными пользователями.

Носителем спам-информации является текст сообщения электронной почты. С лингвистической точки зрения на организацию текста спама будут

оказывать непосредственное влияние две тенденции: соблюдение стилистических рамок электронного письма как жанровой разновидности Интернет-коммуникации и использование психолингвистических приемов построения максимально эффективного рекламного текста. Они находят отражение в структуре электронного сообщения, в его композиции, в выборе лексических, стилистических и прагматических средств, соответствующих намерениям автора сообщения.

Основными чертами спама являются:

- анонимный отправитель (указан вымышленный или сгенерированный автоматически адрес);
- односторонний характер коммуникации, выражающийся в пренебрежении интересами получателя;
- императивный характер значительного числа сообщений (побуждение к приобретению товаров, услуг, посещению веб-страниц и т.д.);
- намеренное нарушение орфографии, которое не препятствует адекватному восприятию сообщения, но требует дополнительных усилий со стороны получателя (используется с целью преодолеть программные фильтры) [1, с. 42].

Каждое из сообщений образовано двумя основными компонентами: темой электронного сообщения и телом письма. Тема электронного сообщения – это микротекст, выполняющий собственную коммуникативную функцию и обладающий определенной степенью автономности по отношению к основному тексту сообщения. Целью этого текста является привлечение внимания получателя, стремление ввести пользователя в заблуждение, побуждение его к восприятию основного текста сообщения. Тело письма несет на себе основную смысловую нагрузку в плане реализации намерений отправителя. При составлении эффективного сообщения необходимо учитывать особенности электронной коммуникации, отличающие ее от традиционного общения. Кроме того, возможности электронного сообщения подразумевают широкое использование графики, аудио- и видеоматериалов, особенностей гипертекста.

Задачу фильтрации спама можно рассматривать как задачу классификации входящего потока электронных сообщений на категории «спам» и «не спам». Для решения задачи классификации широкое применение получили нейронной сети, выступающие 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. Их использование для решения задачи классификации состоит в указании принадлежности входного образа, представленного вектором входных признаков одному или нескольким заранее определенным классам.

Применение нейросетевой технологии предусматривает выполнение следующих основных этапов [4]:

1) выбор структуры сети (задание входных, выходных параметров сети, определение числа ее слоев и нейронов в каждом слое);

2) обучение нейронной сети выбранного типа на данных, сформированных из базы электронных почтовых сообщений;

3) применение обученной нейронной сети для классификации новых почтовых сообщений на категории «спам» / «не спам».

Особенность использования обученной нейронной сети для решения поставленной задачи определяется ее обобщающей способностью, которая заключается в возможности точно классифицировать не только ранее выявленные спамовые электронные почтовые сообщения, но и распознавать новые виды спама. Веса обученной нейронной сети хранят достаточное количество информации о спамовых письмах, что определяет эффективность применения данной технологии.

Непосредственное построение эффективной нейросетевой модели спам-фильтрации возможно в рамках технологии обнаружения знаний в базах данных, включающей следующие этапы [2, с. 12]:

1) получение исходных данных электронных почтовых сообщений, включающих примеры спамовых и неспамовых писем;

2) предварительная обработка исходных данных и формирование обучающей выборки для обучения нейронной сети;

3) разработка структуры нейронной сети: задание входов, выходов, числа слоев сети и нейронов в каждом слое;

4) обучение сети для построения модели спам-фильтрации;

5) тестирование и оценка нейросетевой модели спам-фильтрации.

Поскольку исходные письма представляют собой тексты в электронном виде, необходимо из исходной текстовой информации предварительно выделить значимые параметры для анализа. Другими словами, необходимо выработать четкий набор параметров, характеризующих электронные почтовые сообщения и позволяющих производить их классификацию по категориям «спам»/«не спам». Значения выделенных параметров затем войдут в обучающую выборку. Далее необходимо создать набор данных из различных источников, на основании которого будет строиться решение поставленной задачи. Полученные исходные данные представлены в табличном виде, где каждая строка соответствует отдельному письму, а каждый столбец соответствует отдельному признаку письма. В ячейках таблицы представлены значения признаков, характеризующих конкретное электронное почтовое сообщение.

Таблица с исходными данными является еще сырым материалом для применения методов интеллектуального анализа, поэтому данные, входящие в нее, необходимо предварительно обработать. Во-первых, таблица может содержать параметры, имеющие одинаковые значения для всего столбца. Такие признаки не индивидуализируют исследуемые объекты, следова-

тельно, их надо исключить из анализа. Во-вторых, таблица может содержать некоторый категориальный признак, значения которого во всех записях различны. Очевидно, что это поле нельзя использовать для анализа данных и его надо исключить. Параллельно с очисткой данных по столбцам таблицы также необходимо провести предварительную очистку данных по строкам. Любая база данных обычно содержит ошибки, неточно определенные значения, соответствующие каким-то редким, исключительным ситуациям, и другие дефекты, которые могут снизить эффективность фильтрации спама. Такие записи необходимо отбросить, поскольку даже если подобные «выбросы» не являются ошибками, а представляют собой редкие исключительные ситуации, они все равно вряд ли могут быть использованы, поскольку по нескольким точкам статистически невозможно судить об искомой зависимости в данных.

Анализ статистических признаков нейронной сетью напоминает байесовскую фильтрацию спама, где для каждого слова или словосочетания можно установить коэффициент «спамности». Однако в отличие от байесовского фильтра здесь связи между нейронами способны 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 байесовским 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.

Нейронная сеть способна к самообучению, обнаружению ранее неизвестных спам-сообщений, в то время как эффективность байесовского фильтра зависит от постоянной коррекции коэффициентов на новых выборках, нет процесса самообучения. Для каждого нового спам-сообщения при использовании байесовского 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. Нейронная сеть избавлена от многих недостатков байесовского фильтра, однако эффективность метода зависит от обучающей выборки, используемой в процессе обучения. В итоге возникает задача правильного формирования обучающей выборки, обладающей репрезентативностью и достоверностью. При неудовлетворительных результатах оценки модели необходимо вернуться к одному из этапов и выполнить все последующие этапы в указанной последовательности.

В ходе исследования была создана компьютерная программа, использующая наивный байесовский классификатор, который способен ответить на вопрос, к какой категории классов «спам» – «не спам» относится электронное сообщение. Данная программа написана на языке программирования Python и использует корпус СМС-сообщений в формате CSV в качестве обучающего алгоритма материала.

ЛИТЕРАТУРА

1. *Ажмухамедов, И. М.* Усовершенствованный метод фильтрации нежелательного трафика / И. М. Ажмухамедов, К. В. Запорожец // Вестн. Астрахан. гос. тех. ун-та. Сер. Управление, вычислитель-ная техника и информатика. – 2014. – № 1. – С. 42–47.
2. *Катасёв, А. С.* Разработка нейросетевой системы классификации электронных почтовых сообщений / А. С. Катасёв // Вестн. Казан. гос. энергетич. ун-та. – 2015. – № 1 (25). – С. 12–15.
3. *Кудашова, Н. Н.* Коммуникативная теория текста и текстовый антропоцентризм [Электронный ресурс] / Н. Н. Кудашова // Балт. гуманит. журн. – 2017. – № 3. – Режим доступа : https://cyberleninka.ru/article/n/kommunikativnaya-teoriya-teksta-i-tekstovyy-antropotsentrizm. – Дата доступа : 16.01.2021.
4. *Мироненко, А. Н.* Автоматическая фильтрация спама на базе сети формаль-ных нейронов [Электронный ресурс] / А. Н. Мироненко // Вестн. ОмГУ. – 2011. № 2. – Режим доступа : https://cyberleninka.ru/article/n/avtomaticheskaya-filtratsiya-spama-na-baze-seti-formalnyh-neyronov. – Дата доступа : 12.01.2021.
5. *Солдатова, А. А.* О возможности применения спам-фильтра для анализа текста / А. А. Солдатова // Вестн. Твер. гос. ун-та. Сер. Филология. – 2014. – № 4. – С. 342–346.