Р. В. Детскина, К. П. Антоник

ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ФИЛЬТРАЦИИ СПАМА

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 грамматические, лексические и синтаксические особенности спам-сообщений.

Задачу фильтрации спама можно рассматривать как задачу классификации входящего потока электронных сообщений на категории «спам» и «не спам». Для ее решения широкое применение получили нейронные сети, выступающие в качестве механизма принятия решений, предоставляя на выходе вероятностную оценку спама 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 (М. Е. Сухопаров 2015). 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 (А. Н. Мироненко 2011):

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, заключающейся 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 (А. С. Катасёв 2015):

1) получение исходных данных электронных почтовых сообщений, включающих примеры спамовых и не спамовых 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 (А. С. Катасёв 2015). Такие признаки не индивидуализируют исследуемые объекты, следовательно, 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 они 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.

Нами была создана компьютерная программа, использующая наивный байесовский классификатор, который может ответить на вопрос, к какой

категории классов «спам» / «не спам» относится конкретное электронное сообщение. Данная программа написана на языке программирования Python и в качестве обучающего алгоритма материала использует корпус СМС сообщений в формате CSV.