Применение искусственных иммунных систем в фильтрации спам-писем

С. К. Ганиев, email: sharofidinov1990@gmail.com
Ш. Ж. Хамидов, email: hamidov.sherzod.1990@mail.ru

Ташкентский университет информационных технологий имени Мухаммада ал-Хоразмий

Аннотация. *Вместе с ростом количества пользователей увеличиваются риски в области безопасности электронной почты. Спам электронной почты является одним из основных проблем современного Интернета, приносящий финансовый ущерб компаниям и раздражающий пользователей. Традиционные методы фильтрации спама включают в себя такие проблемы, как низкий уровень обнаружения ошибок, высокий уровень ложных срабатываний, слабая самоадаптация и т.д. Среди подходов, разработанных для борьбы со спамом, важными и популярными являются искусственные иммунные системы. В данной работе рассматривается иммунная система для фильтрации спама.*

Ключевые слова: *интеллектуальные, гибридные, спам-фильтры, фильтрация, извлечение признаков, искусственная иммунная система, антиген, антитело, библиотека генов.*

Введение

Проблема нежелательных электронных сообщений в настоящее время является серьёзным вопросом безопасности электронной почты. Спам - незаконные и негативные явления, которые включают в себя использование сервисов электронной почты для отправки нежелательных писем с вредоносными ссылками и вложениями.

Для фильтрации спама было предложено множество решений, их можно разделить на три категории: простые подходы, интеллектуальные подходы и гибридные подходы. Простые подходы, могут быть легко реализованы, но при этом новые разновидности спама очень трудно обнаружить. Интеллектуальные подходы в последние годы играют всё более важную роль в борьбе со спамом благодаря своей способности к самообучению и хорошей производительности. Проектированная система обнаружения спама на основе одного подхода может быть легко взломана, поэтому для улучшения производительности и преодоления

недостатков каждого отдельного подхода предлагаются гибридные подходы, объединяющие два или более подхода.

Новые подходы к фильтрации спама разрабатываются постоянно, и методы, используемые для оценки этих новых подходов и анализа их эффективности, должны идти параллельно со временем.

1. Фильтрация спам-писем

Спам-фильтры играют важную роль в обеспечении безопасности систем электронной почты, имеют определённые метрики и алгоритмы для фильтрации писем. Важным этапом проблемы фильтрации спама является ее распознавание и классификация.

При классификации электронной почты существующие методы борьбы со спамом сталкиваются с двумя основными проблемами. Входящее письмо распознаётся как спам и удаляется, либо спам письмо принимается как не спам. Этот процесс называется ложно положительным и ложно отрицательным соответственно. Ложное срабатывание происходит, когда письма или данные, классифицируются как спам, в то время как письма или данные, которые должны быть удалены, распознаются как не спам.

Важным этапом фильтрации спама является извлечение признаков из электронного письма. Эффективность стратегии извлечения признаков может оказать непосредственное влияние на общие результаты классификации и производительности в зависимости от ее точности, уникальности, надёжности и гибкости. Традиционные методы извлечения признаков:

Методы выбора термина используются для оценки значимости термина или признака, или количества информации, которую содержит термин или признак, для задания классификации, чтобы уменьшить вычислительную сложность и влияние зашумленных терминов или признаков [1].

Извлечение признаков из текста очень важно, оно играет решающую роль в классификации и напрямую влияет на точность. Этот метод извлечения признаков обычно содержит следующие этапы: выбор термина, извлечение признаков и отображение. При этом набор ключевых слов часто используется для текстовых признаков. Признаки электронных писем извлекаются и отображаются в единой форме [2].

Извлечение признаков на основе изображений. Электронные письма иногда содержат изображения с целью рекламы и маркетинга. Существуют различия между изображениями спама и обычными изображениями (атрибуты, цвета, текст, фон и т.д.), и в соответствии со значительными различиями между этими двумя категориями

изображений было предложено извлечение признаков на основе изображений [3].

Извлечение признаков на основе поведения. Различия между спамом и обычными письмами не только по содержанию, но и по цели отправки, способу передачи, диапазону взаимодействия и т.д. Обычно спаммеры принимают определённые меры для обхода спам-фильтров. Таким образом, мы можем отличить спам от нормальных писем, извлекая различные признаки поведения в процессе отправки писем [4].

2. Искусственная иммунная система для фильтрации спама

Спам приходит в самых разных формах с некоторыми закономерностями. Подобно биологической иммунной системе, полное сообщение электронной почты, как заголовки, так и тело, рассматривается как антиген, а антителами являются цифровые биты шаблонов, которые описываются как регулярные выражения, чтобы приблизительно соответствовать антигенам. В результате, иммунная система может классифицировать сообщения посредством идентификации шаблонов. Существует несколько элементов, определённых для иммунной системы спама: антитело используется в качестве детектора, библиотека генов используется для создания антител, классифицированные сообщения используются для обучения антител, а входное сообщение просто должно быть сопоставлено для завершения фильтрации спама.

Представление детекторов: основным компонентом иммунной системы спама являются детекторы, которые называются антителами. Антитела необходимы для проверки содержимого сообщения электронной почты. Антитела могут быть описаны векторами атрибутов. Чтобы преобразовать сообщение электронной почты в антитело, сначала вычисляется локальный вес каждого слова в сообщении.

Взвешенное сходство: механизм сходства основан на локальных весах, а не на подсчёте схожих слов в двух электронных письмах. Следующим шагом является проверка этого сходства с пороговым значением, чтобы указать класс электронной почты.

Динамическое определение порога: в модели не используется статическое значение порога. Вместо этого используется пороговая функция, напрямую зависящая от количества схожих слов в обоих проверенных письмах.

Библиотека генов: слова, извлечённые из обучающих элементов и ячеек памяти, хранятся в библиотеке генов. Эта библиотека используется для выполнения мутации. Слово из этой библиотеки

заменяет слово из вектора признаков ячейки, как будет описано в алгоритме.

Кэширование антител: слова, имеющие наибольший вес из генной библиотеки, кэшируются в буферном слое антител.

Сортировка детекторов: набор детекторов сортируется после тестирования. Такая сортировка повышает производительность процесса фильтрации.

Отзыв пользователя: классифицированные спам письма не удаляются. Вместо этого система сохраняет его во временной папке. Если пользователь удалил электронное письмо из этой временной папки, это означает, что система провела правильную классификацию [5].

Алгоритм спам-фильтра на основе искусственной иммунной системы. Первым шагом является обучение детекторов. Как упоминалось ранее, в процессе обучения система использует как спам, так и не спам письма. После обучения каждое антитело, существующее в результате обучения, представляет собой пример заранее определенного спамового письма. Кэшированный слой антител строится из генной библиотеки на основе, имеющих наибольший вес в библиотеке генов. Новое поступающее электронное письмо, которое необходимо классифицировать, называется антигеном. Для проверки антигена он сначала подвергается предварительной обработке в соответствии с определением детекторов, рассмотренным ранее. Антиген будет представлен всем антителам, после чего рассчитывается сходство. Если значение сходства превышает пороговое значение, оно классифицируется как спам, в противном случае оно классифицируется как не спам и пропускается в папку входящих сообщений пользователя.

Возможны различные варианты:

- если антиген классифицирован как не спам, а пользователь переместил его в папку спам, это классификация считается ложной, система добавляет этот антиген к антителам в обучающем наборе как спам;
- если антиген классифицирован как спам, он будет перемещен в папку спама;
- если пользователь удалил антиген из папки спама, он подтверждает, что это письмо представляет собой спам, что означает, что классификация была верной;
- если пользователь переместил антиген в свой почтовый ящик, то процесс классификации считается ложным, а антитела, распознавшие этот антиген, удаляются из обучающего набора.

На рис. **Ошибка! Источник ссылки не найден.** показана модель искусственной иммунной системы для фильтрации спама.

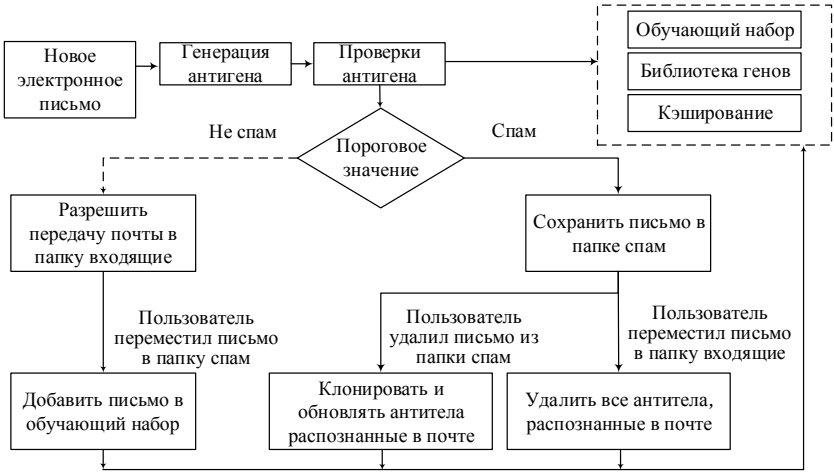

Рис. 1. Модель искусственной иммунной системы для фильтрации спама

Заключение

В современных методах борьбы со спамом интеллектуальные методы обнаружения являются наиболее эффективными и перспективными подходами. Извлечение признаков является основной частью интеллектуальной системы обнаружения спама, которая играет решающую роль в эффективности классификации. В данной работе были рассмотрены этапы фильтрации спам писем, способы извлечения признаков и модель искусственной иммунной системы для фильтрации спама. Комбинирование искусственной иммунной системы с классическими статистическими методами может эффективно улучшить работу спам-фильтров.

Список литературы

1. Koprinska, I. Learning to classify e-mail / I.Koprinska and others // Information Sciences – 2007. – №10. – P. 2167–2187.

2. Oda, T. Developing an immunity to spam / T. Oda, T. White // In Genetic and Evolutionary Computation GECCO 2003 Springer, New York – 2003. – P. 231–242.

3. Gao, Y. A comprehensive approach to image spam detection: From server to client solution / Y. Gao, A. Choudhary, G. Hua // IEEE Transactions on Information Forensics and Security – IEEE, Evanston, IL – 2010. – №5(4). – P. 826–836.

4. Chi-Yuan, Yeh. Effective spam classification based on meta-heuristics / Yeh Chi-Yuan, Wu Chih-Hung, Doong Shing-Hwang // In Systems, Man and Cybernetics, 2005 IEEE International Conference – IEEE, Waikoloa, HI, – 2005. – №4. – P. 3872–3877.

5. Haggag, M. H. Artificial immune system for spam filtering / Haggag M. H., Fattoh E. // International Journal of Intelligent Computing and Information Sciences (IJICIS) – 2009. – №9(2). – P. 117–129.