УДК 004.738.5
К.Т. Кузьма, к.т.н.
ВНЗ «Миколаївський політехнічний інститут», Україна

АНАЛІЗ МЕТОДІВ ФІЛЬТРАЦІЇ СПАМУ

K.T. Kuzma, Ph.D.

ANALYSIS OF METHODS OF SPAM FILTERING

Одним з напрямків досліджень у галузі інформаційних технологій захисту інформації є розробка методів і алгоритмів фільтрації потоку електронної пошти. Разом з розвитком електронної пошти збільшується і кількість загроз її нормальному функціонуванню. Найбільш серйозною і важливою проблемою став SPAM (спам) - небажана комерційна електронна пошта, відправлена без законних підстав великій кількості адресатів.

Аналіз, розробку методів розпізнавання спаму досліджено в роботах О. М. Певзнера, М.О. Семенової, А.М. Мироненка, Г. Робінсона [1-4].

Комплексний захист від спаму складається з наступних етапів:

- аналіз відправника;
- використання фільтрів;
- аналіз змісту листа.

Технічно дані етапи базуються на двох основних підходах фільтрації SPAM – фільтрація за формальними ознаками повідомлення (за способом посилки й оформленню) і за його змістом (семантичні методи фільтрації).

Формальні методи включають фільтрацію за списками (поштових адрес, IP-адрес) та за формальними ознаками листа (наявність багатьох відправників, відсутність одержувача, формат, розмір тощо).

Семантичні методи передбачають розпізнавання за змістом листа (словосполучення, евристики, статистика) або розпізнавання за зразками листів (за сигнатурами).

Для роботи семантичних методів використовуються фільтри здатні до самонавчання, при цьому необхідно здійснювати їх початкове навчання, тобто розсортовувати вручну листи для виявлення статистичних особливостей нормальних листів і SPAM. Таким чином, задача фільтрації SPAM, розглядається як задача класифікації - визначення належності об'єкта (електронного повідомлення) до одного з заздалегідь виділених класів (спам і «не спам») на підставі аналізу сукупності ознак, що характеризують даний об'єкт.

В основі фільтра лежить список ознак, за якими проводиться аналіз повідомлення і обчислюється умовна ймовірність спамності за кожного ознакою. Загальна ймовірність спаму повідомлення визначається за одним з методів:

1. Об'єднуються всі ймовірності за теоремою Байєса.
2. Ймовірності комбінуються і перевіряються на скільки отримана множина схожа з випадковою (метод Фішера).

Теорема Байєса лежить в основі багатьох сучасних систем штучного інтелекту, призначених для роботи в умовах невизначеності. Такі системи дають ймовірнісну оцінку, тому звичайно не заміняють експерта, а забезпечують підтримку прийняття рішення.

Нехай $F_S(W_i)$ – кількість SPAM-листів, у яких зустрілося слово W_i, а $F_{NS}(W_i)$ – кількість корисних листів, у яких зустрілося слово W_i; H_S – гіпотеза про те, що лист є SPAMом, H_{NS} – корисний лист. Тоді ймовірність того, що поява слова W_i у листі означає SPAM, обчислюється за формулою:

$$P(W_i|H_s) = \frac{F_s(W_i)}{F_s(W_i) + F_{NS}(W_i)},$$

а ймовірність того, що слово W_i не вказує на SPAM у листі:

$$P(W_i|H_{NS}) = \frac{F_{NS}(W_i)}{F_s(W_i) + F_{NS}(W_i)}.$$

Якщо вектор W включає всі m слів нового листа, то ймовірність того, що він SPAM, обчислюється за формулою Байєса таким чином:

$$P(H_S|W) = \frac{\prod_{j=1}^{m} P(W_j | H_s)}{\prod_{j=1}^{m} P(W_j | H_s) + \prod_{j=1}^{m} P(W_j | H_{NS})}.$$

Віднесення листа до SPAMу або корисних листів виконується з врахуванням заданого програмістом, адміністратором, користувачем почтової програми спам-фільтрації значення ймовірності, яке становить 0,6 - 0,8. Після ухвалення рішення щодо листа в базі даних обновляються імовірнісні бази для слів, які входять до нього.

Максимальний результат, досягнутий байєсовськими фільтрами на сьогоднішній день складає близько 95 % відфільтрованого спаму. Існують безліч модифікацій, які дозволяють збільшити ефективність фільтра: метод градуйованої фільтрації «спаму» дозволяє підвищити якість оцінки даних за рахунок врахування наступних параметрів - кількості листів, в яких зустрічалися слова певної категорії, частоти використання слів у листах певної категорії, використання слів, що вперше зустрілися в листі і не існували до цього в базі [2]; побудова фільтра на основі багатошарового персептрона, що дозволяє враховувати семантичні зв'язки автоматично [3]. Перевага нейромережевого підходу перед байєсівським полягає в тому, що не робиться ніяких попередніх припущень щодо характеру небажаних повідомлень, а семантичні зв'язки враховуються автоматично.

Таким чином, розвиток нейромережевого підходу фільтрації небажаних повідомлень забезпечить можливість створення прикладних систем індивідуального захисту від небажаної кореспонденції для персональних комп'ютерів з використанням технологій штучного інтелекту.

Література

1. Певзнер О. М. Моделювання та аналіз ефективності зниження спам-ризиків за допомогою марківської фільтрації // Матеріали VI Всеукраїнської науково-практичної конференції «Комп'ютерне моделювання та інформаційні технології в науці, економіці та освіті». – Кривий Ріг, 26–28 квітня 2005 р. – С. 156–164.

2. Семенова М.А. Модель и метод градуированной фильтрации «СПАМА» : автореф. дис. на соискание науч. степени канд. техн. наук: спец. 05.13.19 "Методы и системы защиты информации, информационная безопасность" / М.А. Семенова – Санкт-Петербург, 2009. – 20 с.

3. Мироненко А.Н. Алгоритм контентной фильтрации спама на базе совмещения метода опорных векторов и нейронных сетей : автореф. дис. на соискание науч. степени канд. техн. наук: спец. 05.13.19 "Методы и системы защиты информации, информационная безопасность" / А.Н. Мироненко – Санкт-Петербург, 2012. – 18 с.

4. Robinson G. A Statistical Approach to the Spam Problem / G/ Robinson // Linux Journal, 2003. - Issue #107. - Режим доступу: http://www.linuxjournal.com/article/6467.