ПОСТРОЕНИЕ СПАМ-ФИЛЬТРА НА ОСНОВЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

А.А. Григорьев, А.В. Галковский, Д.С. Совпель, Д.А. Клебанов

Безопасность и качество информации, поступающей из интернета один из ключевых вопросов его развития. Количество информации экспоненциально увеличивается каждый день и вопрос в том, как ее обрабатывать становится достаточно острым. Одним из подходов для анализа больших данных является алгоритмы машинного обучения. Машинное обучения применяется в том числе для фильтрации данных и защиты пользователя от нежелательного контента.

Существует множество подходов к построению алгоритмов для фильтрации спама:

– фильтры содержимого: анализ содержимого сообщений, поиск по словам, которые обычно используются в спам-письмах;

– фильтры черного списка: игнорирование электронных писем, приходящих с IP-адресов и e-mail адресов, находящихся в черных списках (некоторые фильтры также могут проверять IP-репутацию IP-адреса);

– фильтры на основе правил: применяют настраиваемые правила, разработанные организацией, чтобы исключить электронные письма от определенных отправителей или электронные письма, содержащие определенные слова в строке или тексте темы.

Наиболее эффективным является подход комбинированных систем, основным компонентом которого является фильтр содержимого, основанный на алгоритмах машинного обучения. Широко применяемым алгоритмом фильтрации спама является алгоритм наивного Байеса. Он возвращает вероятность того, что сообщение является спамом или не спамом при условии исходных данных. Расчет вероятности производится при помощи формулы Байеса [1]:

Таким образом при помощи этого алгоритма и набора исходных данных, представляющих размеченные на спам и не спам письма, можно построить систему спам фильтрации и использовать ее в системе почтовых клиентов.

Литература

1. Bishop C. Pattern Recognition and Machine Learning. – Springer: 2006. – 48 p.