

ОБЗОР СИСТЕМ ФИЛЬТРАЦИИ ЗАПРЕЩЕННОГО ИНТЕРНЕТ-КОНТЕНТА

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Зычков С. А.

Самаль Д. И. – к. т. н., доцент

В современном мире становится все больше информации в открытом доступе. Возникает необходимость ограждать себя от нежелательной информации, для чего могут служить фильтры интернет-контента.

Фильтры контента могут защищать от нежелательной информации различных типов: текста, изображений, аудио и видео. Большинство систем фильтрации интернет-контента основано на базах ссылок, в которые они были включены из-за содержания. В связи с медленной скоростью обновления баз и блокирования всего контента находящегося по ссылке, данный тип фильтрации является ненадежным и малоэффективным.

Для повышения эффективности создаются фильтры с динамическим анализом информации на стороне клиента или прокси. Они анализируют конкретную информацию (видео, звук, изображения или текст) и определяют принадлежность информации к разрешенной или запрещенной.

Примером может служить фильтр "Licenzero" имеющий функционал определения материалов порнографического содержания в видео. Для определения видеоролика к классу запрещенного контента используются несколько алгоритмов, которые анализируют видео не целиком, а фрагменты из него. Используемые алгоритмы ищут в видео 4 типа признаков: характер движений (ритмичность), цвет, содержание кадра (поиск по определенным шаблонам) и используемый звук. По каждому алгоритму высчитывается своя вероятность, что видео относится к запрещенному. И после этого высчитывается общая вероятность.

Для определения характера движения используются алгоритмы с использованием машинного обучения. Сначала система обучается на тестовых наборах, а потом использует фильтры при классификации видео. Использовались методы "AdaBoost" и "Support Vector Machines". Для определения движения используются пространственно-временные фильтры основанные на применении операции свертки сигнала и суммирования. Для классификации определенного движения применяются методы "байесовских спам-фильтров" и метод опорных векторов.

Анализируя цвет кожи, определяем наличие человеческих обнаженных тел в кадрах видео. Для этого можно использовать алгоритм определения человеческой кожи на изображении "skin detection".

При анализе звука уделяется внимание двум вещам: наличие человеческого голоса и ритмичное повторение одних и тех же звуков. Для этого используется mel-частотные кепстральные коэффициенты.

Основываясь на суммарной вероятности можно классифицировать видео к запрещенному контенту и заблокировать его.

Недостатком динамических систем является точность классификации. У рассмотренной системы при тестах в оптимальных условиях вероятность верной классификации составила 78,3%. Несмотря на большой процент ошибок динамические системы определения запрещенного контента являются более эффективными по сравнению с базами сайтов, содержащих запрещенный контент.

Для уменьшения количества ошибок можно использовать другие алгоритмы определения, разрешить проводить обучение системы пользователям (появится возможность персонального определения понятия запрещенного контента), а также при ошибках – возможность ручной правки в классификации, для понижения количества ошибок при следующем анализе.

Список использованных источников:

1. Контент-фильтр для Linux и Windows / Электронный доступ // <http://habrahabr.ru/sandbox/36893/>
2. Licenzero / Электронный доступ // <http://habrahabr.ru/post/116173/>