БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Факультет радиофизики и компьютерных технологий
Кафедра интеллектуальных систем

Аннотация к магистерской диссертации

«Фильтрация спам изображений с использованием методов машинного обучения»

специальность 1-98 80 01 «Методы и системы защиты информации, информационная безопасность»

Радкевич Артём Сергеевич

Научный руководитель: кандидат физико-математических наук, доцент Козадаев Константин Владимирович

2018

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Магистерская диссертация: 4 главы, 106 страниц, 115 рисунков, 10 таблиц, 27 использованных источников, 3 приложения.

МАШИННОЕ ОБУЧЕНИЕ, ПРИЗНАК, СПАМ, ИЗОБРАЖЕНИЕ, ОЦЕНКА, МЕТОД, АЛГОРИТМ, ИНФОРМАТИВНОСТЬ, ТОЧНОСТЬ, SVM, PCA, EIGENFACES, AUC, ГРАФИК, ВЕКТОР ВЕСОВ.

По причине быстрого развития сетевых платформ и средств для коммуникации между людьми так же увеличилось количество вредоносного контента, частью которого являются спам изображения.

Цель работы заключалась в исследовании метода обнаружения спам изображений только на основе свойств самих изображений и в оценке эффективности такого подхода.

В качестве объекта исследования были выбраны методы для классификации на основе SVM и PCA. Реализация метода SVN осуществлялась двумя способами. Первый подход осуществлял выборку всех указанных свойств изображений и, на основе полученных признаков, выполнялась классификация изображений. Второй подход реализовывал выборку наиболее информативных признаков, после чего проводил классификацию изображений. Реализация метода PCA основывалась на технологии «eigenfaces», задачей которого являлось построение собственного пространства спам изображений и оценка схожести с ним остальных изображений для проведения классификации.

Рассмотренные способы обнаружения спам изображений показали свою эффективность при низких вычислительных требованиях, что позволит использовать их при создании легковесных клиентских спам фильтров.

Во введении описывается угроза спама, обосновывается актуальность темы. В главе 1 рассматриваются различные виды спама, описываются способы противодействия спаму, а также вводится понятие спам изображений и способы их обнаружения. В главе 2 дается краткое определение машинному обучению, а также рассматриваются, используемые в работе, алгоритмы классификации SVM и PCA. В главе 3 описывается реализация способа классификации спам изображений на основе методов SVM и PCA. В главе 4 демонстрируются результаты исследования и проводится анализ результатов. В заключении дается вывод о проделанной работе.

АГУЛЬНАЯ ХАРАКТАРЫСТЫКА ПРАЦЫ

Магістарская дысертацыя: 4 раздзелы, 106 старонак, 115 малюнкаў, 10 табліц, 27 выкарыстаных крыніц, 3 дадатка.

МАШЫННАЕ НАВУЧАННЕ, ПРЫКМЕТА, СПАМ, МАЛЮНАК, АЦЭНКА, МЕТАД, АЛГАРЫТМ, ІНФАРМАТЫЎНАСЦЬ, ДАКЛАДНАСЦЬ, SVM, PCA, EIGENFACES, AUC, ГРАФІК, ВЕКТАР ВАГАЎ.

Па прычыне хуткага развіцця сеткавых платформаў і сродкаў для камунікацыі паміж людзьмі, павялічылася колькасць шкоднаснага кантэнту, часткай якога з'яўляюцца спам выявы.

Мэта работы складалася ў даследаванні метаду выяўлення спам малюнкаў толькі на аснове уласцівасцяў саміх малюнкаў і ў ацэнцы эфектыўнасці такога падыходу.

У якасці аб'екта даследавання былі выбраны метады для класіфікацыі SVM і PCA. Рэалізацыя метаду SVN ажыццяўлялася двума спосабамі. Першы падыход ажыццяўляў выбарку ўсіх паказаных уласцівасцяў малюнкаў і, на аснове атрыманых прыкмет, выконвалася класіфікацыя малюнкаў. Другі падыход рэалізоўваў выбарку найбольш інфарматыўных прыкмет, пасля чаго праводзіў класіфікацыю малюнкаў. Рэалізацыя метаду PCA грунтавалася на тэхналогіі «eigenfaces», задачай якога з'яўлялася пабудова ўласнай прасторы спам малюнкаў і ацэнка падабенства з ім астатніх малюнкаў для правядзення класіфікацыі.

Разгледжаныя спосабы выяўлення спам малюнкаў паказалі сваю эфектыўнасць пры нізкіх вылічальных патрабаваннях, што дасць магчымасць выкарыстоўваць іх пры стварэнні прастых кліенцкіх спам фільтраў.

Ва ўвядзенні апісваецца пагроза спаму, абгрунтоўваецца актуальнасць тэмы. У раздзеле 1 разглядаюцца розныя віды спаму, апісваюцца спосабы процідзеяння спаму, а таксама ўводзіцца паняцце спам малюнкаў і спосабы іх выяўлення. У раздзеле 2 даецца кароткае вызначэнне машыннай навучанню, а таксама разглядаюцца, якія выкарыстоўваюцца ў рабоце, алгарытмы класіфікацыі SVM і PCA. У раздзеле 3 апісваецца рэалізацыя спосабу класіфікацыі спам малюнкаў на аснове метадаў SVM і PCA. У раздзеле 4 дэманструюцца вынікі даследавання і праводзіцца аналіз вынікаў. У заключэнні даецца выснову аб праведзенай.

GENERAL DESCRIPTION OF WORK

Master thesis: 106 pages, 115 figures, 10 tables, 27 sources, 3 applications.

MACHINE LEARNING, FEATURE, SPAM, IMAGE, SCORE, METHOD, ALGORITHM, INFORMATION RATE, ACCURACY, SVM, PCA, EIGENFACES, AUC, GRAPH, VECTOR OF WEIGHTS.

The amount of malicious content, of which spam is a part, has increased because of fast development of network platforms for communication between people.

The goal of the work is to investigate the method of detecting spam images only based on the properties of the images and to evaluate the effectiveness of that approach.

Methods for image classification based on SVM and PCA were chosen as the subject of the research. The implementation of the SVN method was done in two ways. The first approach selected all features of the images and, based on the obtained features, classified them. The second approach was based on selecting the most informative features, and then performed classification of images. The implementation of the PCA method was based on the "eigenfaces" technology, whose task was to build its eigenspace for spam images and to evaluate the similarity with remaining images for the classification.

The considered approaches for detecting spam images have shown their effectiveness at low computational requirements, which will allow using them when creating lightweight client spam filters.

The threat of spam and the relevance of the topic are described in the introduction. The different types of spam, the ways to counter spam, the concept of spam images and how to detect them are considered in chapter 1. A brief definition of machine learning and the algorithms for classifying SVM and PCA are described in chapter 2. The implementation of the approach for classifying spam images based on SVM and PCA methods is described in chapter 3. The results of the study and an analysis of them are demonstrated chapter 4. The findings of the performed research are given in conclusion.