Министерство образования и науки Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра Математического и компьютерного моделирования

наименование кафедры

Искусственный интеллект в компьютерной безопасности

название темы выпускной квалификационной работы

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студентки 4 курса 451 Группы

направления 38.03.05 – Бизнес информатика

код и наименование направления

механико-математического факультета

наименование факультета, института, колледжа

Крыловой Марии Валериевны

фамилия, имя, отчество

Научный руководитель

доцент, к.ф.-м.н., доцент		С. П. Шевырев
должность, уч. степень, уч. звание	подпись, дата	инициалы, фамилия

Зав. кафедрой

зав. каф., д.ф.-м.н.		Ю. А. Блинков
должность, уч. степень, уч. звание	подпись, дата	инициалы, фамилия

Саратов 2017

ВВЕДЕНИЕ

Актуальность темы исследования. История искусственного интеллекта (ИИ) начинается задолго до нашей эры. Аристотель был первым, кто попытался определить законы «правильного мышления» или процессы неопровержимых рассуждений. Попытки создания механических счетных устройств в средние века сильно впечатляли современников. Наиболее известна машина Паскалина, построенная в 1642 г. Блезом Паскалем. Паскаль писал, что «арифметическая машина производит эффект, который кажется более близким к мышлению по сравнению с любыми действиями животных».

Возможности же практической реализации ИИ появились с момента создания электронных вычислительных машин. В это время развернулась философская дискуссия на тему «Может ли машина мыслить?». Итогом этой дискуссии стал тест, предложенный Аланом Тьюрингом в 50-е гг. XX века [1]. Тест заключается в следующем: Имеются два телетайпа (в то время других терминальных устройств не было, сейчас бы предложили ICQ). Один из телетайпов подключен к машине, другой — к аппарату, за которым сидит человек. Несколько экспертов поочередно ведут диалог на каждом из телетайпов. Если большинство экспертов не смогут в течение пяти минут распознать в одном из собеседников машину, то тест Тьюринга считается пройденным успешно.

Тест Тьюринга сыграл определенную роль в развитии искусственного интеллекта, в том числе и критика самого теста. Здесь можно провести аналогию с авиацией. Хорошими летательными аппаратами, по логике теста Тьюринга, должны считаться такие, которые неотличимы от птиц до такой степени, что даже птицы принимают их «за своих». Развитие авиации началось тогда, когда конструкторы перестали копировать птиц, а занялись аэродинамикой, материаловедением и теорией прочности. Робототехника стала индустрией после того, как перестала копировать анатомию человека.

Аналогично, субъекты искусственного интеллекта получили право на жизнь после того, как прекратились попытки построить системы ИИ, думающие и действующие подобно людям, а начали строить системы, действующие и думающие рационально, т.е. достигающие наилучшего результата.

Актуальность данной темы заключается в важности развития технологии искусственного интеллекта для таких прогрессивных, на сегодняшний день, отраслей науки как робототехника, кибернетика и для более быстрого, удобного доступа к мировым информационным ресурсам. А также искусственный интеллект применяется для развития социальных услуг и для обеспечения компьютерной безопасности, в частности для защиты от спама.

Актуальность определила выбор **темы** данной работы: «Искусственный интеллект в компьютерной безопасности»

Цель работы показать практически, как искусственный интеллект применяется для классификации электронных писем на «спам» и «неспам».

Объект исследования — электронные письма.

Предмет исследования — применение искусственного интеллекта для анализа электронных писем на предмет спама.

Для достижения поставленных целей в работе необходимо решить следующие **задачи**:

1) рассмотреть основные понятия искусственного интеллекта;
2) изучить основные проблемы компьютерной безопасности;
3) раскрыть понятие спама;
4) исследовать применение искусственного интеллекта для анализа электронных писем.

Теоретико-методологической основной исследования явились концепции, раскрывающие сущность искусственного интеллекта и рекомендации по его использованию (Берзон И.И., Буянова Е.А., Кожевников М.А., Чаленко А.В., Галанова В.А., Басова А.И.).

Для решения поставленных задач были использованы следующие теоретические методы исследования: изучение источников, теоретический анализ, обобщение литературных данных, математическая обработка данных.

Научная новизна исследования заключается в следующем:

1) рассмотрены новые направления применения искусственного интеллекта;

2) проанализирован байесовский подход для анализа электронных писем.

Практическая значимость проводимой работы заключается в представлении решения практической задачи, а именно, написании программного кода и представлении результатов графически.

Основное содержание работы. Выпускная квалификационная работа состоит из введения, четырех теоретических и одной практической глав, заключения, списка использованных источников и приложения.

Введение содержит основные положения: статистически подкреплённую актуальность темы исследования; цель, объект, предмет, задачи исследования; практическую значимость исследования.

Первая глава «Искусственный интеллект» содержит основные понятия, историю зарождения искусственного интеллекта; основные подходы к пониманию проблемы искусственного интеллекта; некоторые модели и методы исследования; современные достижения в данной области.

Существует множество определений понятия «искусственный интеллект». Приведем некоторые из них.

Искусственный интеллект —

1. наука и технология создания интеллектуальных машин, особенно интеллектуальных компьютерных программ;
2. свойство интеллектуальных систем выполнять творческие функции, которые традиционно считаются прерогативой человека.

Как прикладная наука «Искусственный интеллект» имеет теоретическую и экспериментальную части. Практически, проблема создания «Искусственного интеллекта» находится на стыке информатики и вычислительной техники — с одной стороны, с нейрофизиологией, когнитивной и поведенческой психологией — с другой стороны. Теоретической основой должна служить Философия искусственного интеллекта, но только с появлением значимых результатов теория приобретёт самостоятельное значение. Пока, теорию и практику «Искусственного интеллекта» следует отличать от математических, алгоритмических, робототехнических, физиологических и прочих теоретических дисциплин и экспериментальных методик, имеющих самостоятельное значение.

На саму возможность мыслить о понятии «Искусственный интеллект» огромное влияние оказало рождение механистического материализма, которое начинается с работы Рене Декарта «Рассуждение о методе» (1637) и сразу вслед за этим работы Томаса Гоббса «Человеческая природа» (1640).

Единого ответа на вопрос, чем занимается искусственный интеллект, не существует. Почти каждый автор, пишущий книгу об ИИ, отталкивается в ней от какого-либо определения, рассматривая в его свете достижения этой науки.

В философии не решён вопрос о природе и статусе человеческого интеллекта. Нет и точного критерия достижения компьютерами «разумности», хотя на заре искусственного интеллекта был предложен ряд гипотез, например, тест Тьюринга или гипотеза Ньюэлла — Саймона.

Можно выделить два направления развития ИИ:

- решение проблем, связанных с приближением специализированных систем ИИ к возможностям человека, и их интеграции, которая реализована природой человека;
- создание искусственного разума, представляющего интеграцию уже созданных систем ИИ в единую систему, способную решать проблемы человечества.

Но в настоящий момент в области искусственного интеллекта наблюдается вовлечение многих предметных областей, имеющих скорее практическое отношение к ИИ, а не фундаментальное. Многие подходы были опробованы, но к возникновению искусственного разума ни одна исследовательская группа пока так и не подошла.

Во второй главе «Компьютерная безопасность» рассматриваются определение знаний в этой области и основные проблемы.

Компьютерная безопасность — меры безопасности, применяемые для защиты вычислительных устройств (компьютеры, смартфоны и другие), а также компьютерных сетей (частных и публичных сетей, включая Интернет). Поле деятельности системных администраторов охватывает все процессы и механизмы, с помощью которых цифровое оборудование, информационное поле и услуги защищаются от случайного или несанкционированного доступа, изменения или уничтожения данных, и приобретает всё большее значение в связи с растущей зависимостью от компьютерных систем в развитом сообществе.

Кибербезопасность — процесс использования мер безопасности для обеспечения конфиденциальности, целостности и доступности данных. Системный администратор обеспечивает защиту активов, включая данные локальной сети компьютеров, серверов. Кроме того, под охрану берутся непосредственно здания и, самое главное, персонал. Целью обеспечения кибербезопасности является защита данных (как в процессе передачи и/или обмена так и находящихся на хранении). В целях обеспечения безопасности данных могут быть применены и контрмеры. Некоторые из этих мер включают (но не ограничиваются) контроль доступа, обучение персонала, аудит и отчётность, оценку вероятных рисков, тестирование на проникновение и требование авторизации.

В третьей главе «Спам» рассмотрены происхождение термина, основные виды спама и способы его распространения.

Спам — массовая рассылка коммерческой и иной рекламы или подобных коммерческих видов сообщений лицам, не выражавшим желания их получать. Распространителей спама называют спамерами.

В общепринятом значении термин «спам» в русском языке впервые стал употребляться применительно к рассылке электронных писем. Незапрошенные сообщения в системах мгновенного обмена сообщениями (например, ICQ) носят название SPIM.

Доля спама в мировом почтовом трафике составляет от 60% (2006) до 80% (2011).

В конце XIX века Western Union позволил отправку телеграфных сообщений в своей сети в многократные места назначения. Первый зарегистрированный случай массовой незапрашиваемой коммерческой телеграммы произошел в мае 1864 года, когда некоторые британские политики получили незапрашиваемую телеграмму, рекламирующую стоматологические

услуги.

Первоначально слово «SPAM» появилось в 1936 г. Оно расшифровывалось как SPiced hAM (острая ветчина) и было товарным знаком для мясных консервов компании Hormel Foods Corporation — острого колбасного фарша из свинины.

Всемирную известность в применении к назойливой рекламе термин «SPAM» получил благодаря знаменитому скетчу «Спам» из известного телевизионного шоу «Летающий цирк Монти Пайтона» (1969) комик-группы из Великобритании «Монти Пайтон».

В 1986 г. в конференциях Usenet появилось множество одинаковых сообщений от некоего Дэйва Родеса, который рекламировал новую финансовую пирамиду. Заголовок гласил: «Заработай кучу денег», а в письмах содержалась инструкция, как это сделать. Автор с завидным упорством продолжал дублировать свои тексты, и они настолько приелись подписчикам, что их стали сравнивать с рекламируемыми в скетче консервами.

Так за словом «спам» закрепилось новое значение, позднее перешедшее в компьютерную терминологию для обозначения назойливых рекламных рассылок.

По данным Лаборатории Касперского, в феврале 2010 года почтовый спам в интернете распределился по тематике следующим образом: в 18,9 % — образование, 15,7 % — отдых и путешествия, 15,5 % — медикаменты, товары/услуги для здоровья, 9,2 % — компьютерное мошенничество, 6,5 % — компьютеры и интернет, 5,2 % — реплики элитных товаров, 4,1 % — реклама спамерских услуг, 2,7 % — для взрослых, 2,2 % — недвижимость, 2,2 % — юридические услуги, 1,9 % — личные финансы, 1,4 % — полиграфия.

Самый большой поток спама распространяется через электронную почту

(e-mail). В настоящее время доля вирусов и спама в общем трафике электронной почты составляет по разным оценкам от 70 до 95 процентов.

Четвертая глава «Использование наивной байесовской модели для классификации» включает в себя теоретические знания о распределении Бернулли, применении наивной байесовской модели для классификации, а также о ее обучении.

Якоб Бернулли - швейцарский математик, один из основателей теории вероятностей и математического анализа. В честь этого ученого была названа группа испытаний, которые моделируются бинарной случайной величиной , для которой вероятность успеха фиксирована и одинакова в каждом независимом испытании. Над такой случайной величиной можно построить множество различных распределений вероятности.

Распределение Бернулли относится к булевым, или бинарным, событиям с двумя возможными исходами: успех (1) и неудача (0). У распределения Бернулли имеется единственный параметр θ, определяющий вероятность успеха: $P(X = 1) = \theta$ и $P(X = 0) = 1 - \theta$. Математическое ожидание распределения Бернулли $E[X] = \theta$, а дисперсия $E[(X - E[X])^2] = \theta(1 - \theta)$.

Обучение вероятностной модели обычно сводится к оцениванию параметров, используемых в этой модели распределений. Параметр распределения Бернулли можно оценить, подсчитав число успехов d в n испытаниях и положив $\hat{\theta} = d/n$. Иными словами, для каждого класса подсчитать, сколько сообщений содержит рассматриваемое слово. Такие оценки на основе относительной частоты обычно сглаживаются путем введения псевдосчетчиков, представляющих исходы виртуальных испытаний с некоторыми фиксированными распределениями. В случае распределения Бернулли в качестве операции сглаживания чаще всего берут поправку Лапласа, которая подразумевает два виртуальных испытания: одно успешное, другое неудачное. Следовательно, оценка на основе относительной частоты заменяется на $(\mathrm{d} + 1)/(\mathrm{n} + 2)$. С точки зрения байесовской модели, это

сводится к принятию равномерного априорного распределения, подразумевающего, что успех и неудача равновероятны. Если того требует ситуация, можно усилить влияние априорного знания, включив большее количество виртуальных испытаний, что означает, что для отодвигания оценки от априорной потребуется больше данных. В случае категориального распределения сглаживание добавляет один псевдосчетчик для каждой из k категорий, что приводит к сглаженной оценке $(\mathrm{d}+1)/(n+k)$. m- оценка представляет собой дальнейшее обобщение, считая параметрами как общее число псевдосчетчиков m, так их распределение по категориям. Оценка для i-й категории определяется как $(\mathrm{d}+\mathrm{p_i}\,\mathrm{m})/(\mathrm{n}+\mathrm{m})$, где $\mathrm{p_i}$ - распределение по категориям (то есть $\sum_{i=1}^{k} p_i = 1$). Отметим, что оценки на основе сглаженной относительной частоты - а значит, и произведение таких оценок никогда не могут достигать крайних значений $\hat{\theta} = 0$ и $\hat{\theta} = 1$.

Пятая глава «Реализация наивной байесовской фильтрации спама» содержит описание наивного байесовского классификатора и программную реализацию фильтрации спама на основе данного классификатора.

Программные спам-фильтры, построенные на принципах наивного байесовского классификатора, делают «наивное» предположение о том, что события, соответствующие наличию того или иного слова в электронном письме или сообщении, являются независимыми по отношению друг к другу. Это упрощение в общем случае является неверным для естественных языков — таких, как английский, где вероятность обнаружения прилагательного повышается при наличии, к примеру, существительного. Исходя из такого «наивного» предположения, для решения задачи классификации сообщений лишь на 2 класса: S (спам) и H=¬S («хэм», то есть не спам) из теоремы Байеса можно вывести следующую формулу оценки вероятности «спамовости» всего сообщения, содержащего слова W_1, W_2, …, W_N:

$$p(S|W_1, W_2, \dots, W_N) = \text{[по теореме Байеса]} = \frac{p(W_1,W_2,\dots,W_N|S)*p(S)}{p(W_1,W_2,\dots,W_N)} =$$

=[так как W_i предполагаются независимыми] = $\frac{\prod_i p(W_i|S)*p(S)}{p(W_1,W_2,\dots,W_N)}$ = [по теореме Байеса] = $\frac{\prod_i \frac{p(W_i|S)*p(S)}{p(S)}*p(S)}{p(W_1,W_2,\dots,W_N)} =$ [по формуле полной вероятности] =

$$= \frac{\prod_i \frac{p(W_i|S)*p(W_i)}{p(S)}*p(S)}{\prod_i\big(p(W_i|S)\big)*p(S)+\prod_i\big(p(W_i|\neg S)\big)*p(\neg S)} =$$

$$= \frac{\prod_i p(S|W_i)*p(W_i)*p(S)^{1-N}}{\prod_i\big(p(S|W_i)*p(S)\big)*p(S)^{1-N}+\prod_i\big(p(\neg S|W_i)\big)*p(W_i)*p(\neg S)^{1-N}} =$$

$$= \frac{\prod_i p(S|W_i)}{\prod_i\big(p(S|W_i)\big)+(\frac{p(\neg S)}{p(S)})^{1-N}*\prod_i\big(p(\neg S|W_i)\big)}$$

Таким образом, предполагая $p(S) = p(\neg S) = 0{,}5$, имеем:

$$p = \frac{p_1 p_2 \dots p_N}{p_1 p_2 \dots p_N + (1-p_1)(1-p_2)\dots(1-p_N)}$$

где:

$p = \Pr(S|W_1, W_2, \dots, W_N)$— вероятность, что сообщение, содержащее слова $W_1, W_2, \dots, W_N$ — спам;

p_1 — условная вероятность $p(S|W_1)$ того, что сообщение — спам, при условии, что оно содержит первое слово (к примеру, «replica»);

p_2 — условная вероятность $p(S|W_2)$ того, что сообщение — спам, при условии, что оно содержит второе слово (к примеру, «watches»);

и т. д.

p_N — условная вероятность $p(S|W_N)$ того, что сообщение — спам, при условии, что оно содержит N-е слово (к примеру, «home»).

Результат p обычно сравнивают с некоторым порогом (например, 0,5),

чтобы решить, является ли сообщение спамом или нет. Если p ниже, чем порог, сообщение рассматривают как вероятный «ham», иначе его рассматривают как вероятный спам.

Для реализации байесовской фильтрации спама использовался язык программирования Python. Для начала были собраны и проанализированные самые популярные спамовые письма. Выявлены наименее значимые слова, которые встречаются как в спаме, так и в «хороших» письмах. Для обучения программы было взято 30 наиболее характерных спамовых писем. Также был составлен список служебных слов русского языка, например, союзы и предлоги, так как они не влияют на результат.

В ходе обучения была получена таблица, представленная на рисунке 1.1, в которой напротив каждого слова дается частота появления данного слова в спаме.

слово	спам	всего
подарок	1	1
даем	1	1
бесплатно	1	1
тестирова...	1	1
системы	4	5
получить	2	3
прямо	1	1
сейчас	1	1
получен	1	1
перевод	3	3
сумму	5	5
для	1	1
одобрена	2	2
выплата	3	4
рублей	5	6
подробнее	2	2
по	2	5
ссылке	1	1
ссылка	1	1
ваш	1	1
запрос	1	1
успешно	1	1
обработан	1	1
выплаты	1	1

слово	спам	всего
ежедневные	1	1
подробно...	6	9
доходность	3	4
неделю	3	4
вывод	3	4
денежных	3	4
средств	3	4
ежедневно	5	6
сайте	6	7
регистрация	2	2
сервере	2	2
вы	2	4
сможете	2	4
получать	4	6
день	3	3
ваши	2	2
первые	2	2
деньги	3	3
получите	4	4
сегодня	5	5
инструкция	2	2
руб	3	5
покажу	2	2
доступ	3	3

слово	спам	всего
заработайте	1	1
сегодня	1	1
проводим	1	1
набор	1	1
сотрудников	1	1
междунар...	1	1
компанию	1	1
платим	1	1
минут	1	1
реальная	1	1
серьезная	1	1
очень	1	1
простая	1	1
работа	1	1
партнера	1	1
которую	1	1
готовы	1	1
платить	1	1
хорошие	1	1
сразу	1	1
скажем	1	1
ничего	1	1
нужно	1	1
регистрир...	1	1

слово	спам	всего
сегодняш...	1	1
средний	1	1
заработок	1	1
сотрудника	1	1
нашей	1	1
компании	1	1
возможно...	1	1
заработать	1	3
базам	1	1
закрыт	1	1
приступить	1	1
работе	1	1
заявка	1	1
рассмотре...	1	1
подтвердите	1	1
её	1	1
регистрац...	1	1
личный	2	3
счет	1	1
системе	2	3
отправлена	1	1
получили	1	2
оплату	1	2
заказу	1	2

переход	1	2
аккаунт	1	2
дата	1	2
время	1	4
сумма	1	2
зачислению	1	2
текущий	1	2
баланс	1	2
выплату	1	2
товаруслуга	1	2
авторизация	1	2
номер	1	2
заказа	1	2

Рисунок 1.1: Частота появления слова в спаме

Программа считает, с какой вероятностью текст, переданный ей на вход, является спамом, в соответствии с рисунком 1.2.

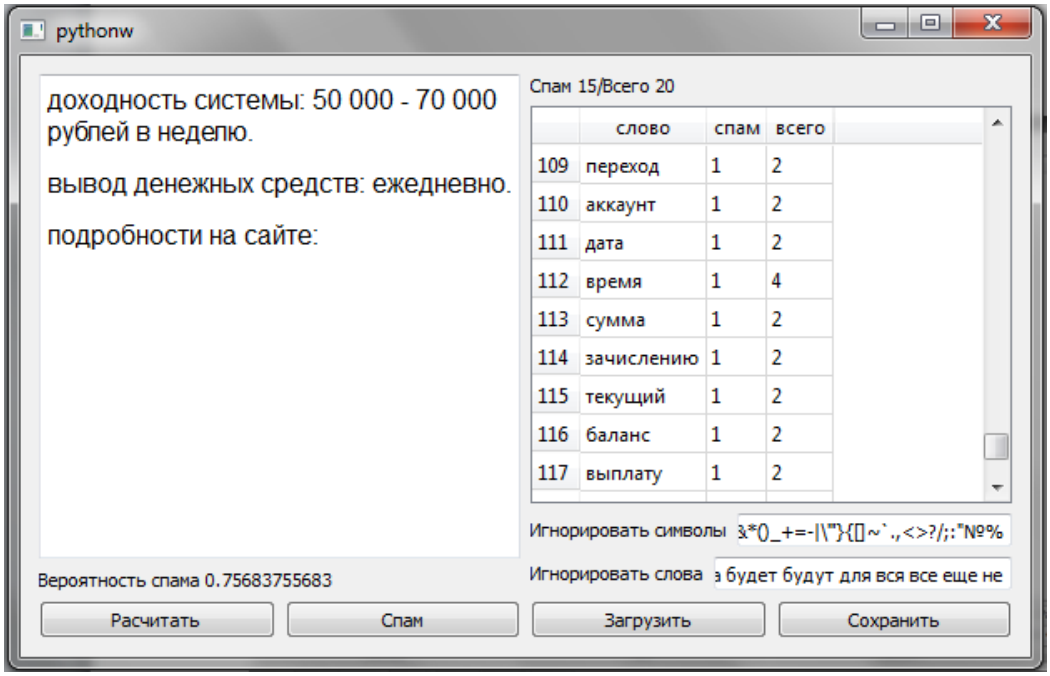

Рисунок 1.2: Окно с примером вычисления вероятности спама

ЗАКЛЮЧЕНИЕ

Трудно представить современную жизнь без искусственного интеллекта, который встречается сегодня чуть ли ни на каждом шагу.

В настоящий момент в области искусственного интеллекта наблюдается вовлечение многих предметных областей, имеющих скорее практическое отношение к ИИ, а не фундаментальное. Многие подходы были опробованы, но к возникновению искусственного разума ни одна исследовательская группа пока так и не подошла.

В ходе выполнения данной работы была достигнута цель исследования, рассмотрено применение искусственного интеллекта для анализа электронных писем.