Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.048

Губский
Михаил Дмитриевич

Система распознавания фальшивых новостей

АВТОРЕФЕРАТ
на соискание академической степени
магистра информатики и вычислительной техники

по специальности 1-40 81 04 – Обработка больших объемов информации

Научный руководитель
Анисимов В.Я.
к.ф-м.н., доцент

Минск 2020

КРАТКОЕ ВВЕДЕНИЕ

Феномен фальшивых новостей известен уже давно. Однако раньше под ним подразумевали дезинформацию, слухи, фальшивые доносы, обман и пр. Но даже в те время фейковые новости считались грозным оружием в политике, экономике и даже в повседневной жизни. К примеру, простой слух о скором бедствии мог поднять спрос до небес на какой-нибудь вид товаров. Стоит упомянуть, что зачастую одного этого оружия хватало, чтобы выиграть битву, а иногда и войну.

В современном информационном мире создавать и распространять фальшивую информацию стало проще. С появлением интернета и социальных сетей каждый может внести неразбериху в сложившийся склад общества. Однако стоит отметить, что не каждая новость является ложной, что только усложняет положение дел.

Важно понимать, что в нынешнее время большинство стран стремятся к демократии и демократическому обществу, что само по себе подразумевает свободу слова и свободу мыслей. Однако, как бы это прекрасно не звучало, данное явление только усугубляет проблему фальшивых новостей. Проблема обостряется, если большинство средств массовой информации являются государственными и, по факту, является предвзятыми к происходящему в стране. В таком случае, правительству не составляет особого труда управлять народом и формировать заведомо ложное мнение в своих целях.

Еще одной особенностью новостей является субъективность восприятия. Подобно слепым мудрецам из притчи о слоне и слепых мудрецах, каждый человек воспринимает факт по своему, даже если он достоверный. При этом большинство людей, услышав новость из какого-то источника, вряд ли будет прилагать усилия, чтобы проверить ее, а скорее поделятся ей с кем-нибудь, другими словами распространят непроверенную информацию.

Помимо социальной безответственности и, порой, недостаточной образованности в некоторых сферах, главными проблемами борьбы с дезинформацией являются ее обнаружение и быстрое распространение. С развитием интернета эти проблемы существенно обострились. В современном мире стало довольно сложно бороться с проблемой распространения ложной информации, поэтому люди сосредоточились на проблеме ее обнаружения.

Чтобы разобраться, что является правдой, а что фейком, люди придумывали всевозможные способы борьбы с дезинформацией. Так в древности использовали письма запечатанные воском с отпечатком герба отправителя, подписи, шифрование и т.п. С переходом информации в цифровой вид некоторые методы стали неприменимы. На их смену пришли цифровые подписи, водяные знаки и др.

На данном этапе развития информационных технологий всё большую популярность набирает сфера машинного обучения. С развитием этого направления появилось множество инструментов для решения задач, с которыми обычный человек не справляется, либо тратит слишком много

времени на их решение. Некоторые из этих инструментов позволяют определить скрытые, порой не видимые для обычного человека признаки и закономерности, которые присущи конкретной специфической задаче. Одним из таких инструментов являются глубокие нейронные сети. Глубокие нейронные сети получили широкое распространение за счет значительного прироста результативности в решении задач регрессии и классификации, после чего их стали использовать и в других задачах машинного обучения.

В данной работе описано исследование методов решения обнаружения фальшивых новостей. Основное внимание уделяется методам машинного обучения и нейронным сетям в частности. Также предложена система, позволяющая просто и удобно вычислять процент подлинности новостного контента.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Целью диссертационной работы является разработка алгоритмов и программного обеспечения для решения задачи обнаружения фальшивых новостей, используя методы машинного обучения и глубоких нейронных сетей в частности.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Собрать набор данных для обучения нейронных сетей.
2. Разработать систему создания и обучения моделей, решающих выбранную проблему.
3. Разработать систему оценки подлинности новости.
4. Провести экспериментальные исследования разработанной системы.

Объектом исследования является новостной контент и его характеристики.

Предметом исследования являются методы и алгоритмы оценки правдоподобности и классификации новостного контента.

Актуальность исследования заключается в том, что проблема, решаемая в данной работе, является острой и насущной. С помощью дезинформации можно легко управлять людьми и подстраивать их мнение под себя, что может пагубно сказаться на каждом человеке и обществе в целом. Простой обыватель не всегда может оперировать достоверными фактами и способствует распространению фальшивой информации. Методы машинного обучения являются передовыми и уже используются повсеместно для решения подобных задач, что говорит об актуальности методов, хорошей поддержке и возможном развитии работы.

Основной *гипотезой*, положенной в основу диссертационной работы, является перспективность использования инструментов машинного обучения и в частности глубоких нейронных сетей при решении задачи классификации новостного контента. Возможности применения данных технологий и полученных результатов способны улучшить качество потребляемой пользователем информации и защитить его от негативного влияния фейков.

Личный вклад соискателя

Результаты, приведенные в диссертации, получены магистрантом лично.

Вклад научного руководителя В. Я. Анисимова, заключается в формулировке целей и задач исследования.

Публикации

По теме диссертации опубликована 1 печатная работа в международном научном журнале.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, четырех глав, заключения, списка использованных источников, списка публикаций автора и приложений. В первой главе представлен теоретический анализ предметной области, выявлены основные существующие проблемы в рамках тематики исследования, показаны направления их решения. Вторая глава посвящена анализу существующих решений с использованием алгоритмов машинного обучения. В третьей главе описана реализация системы создания и обучения моделей для решения поставленной задачи, а также реализация веб-приложения для классификации отдельной новости. В четвертой главе предложены результаты исследования построенных моделей, а также примеры практического применения разработанной системы.

Общий объем работы составляет 71 страницу, из которых основного текста – 40 страниц, 17 рисунков на 14 страницах, 7 таблиц на 3 страницах, список использованных источников из 43 наименований на 4 страницах и 5 приложений на 10 страницах.

ОСНОВНОЕ СОДЕРЖАНИЕ

Во **введении** определена область и указаны основные направления исследования, показана актуальность темы диссертационной работы, дана краткая характеристика исследуемой проблемы, обозначена практическая ценность работы.

В **первой главе** проведен анализ проблемы фальшивых новостей, ее актуальности, а также теоретический методов, используемых журналистами при верификации новостного контента.

В первую очередь решается проблема субъективности восприятия информации путем определения терминов «факт» и «объективность». Данный вопрос является довольно философским, однако, очень важным для дальнейших исследований. Из данных терминов вытекают «верификация» и «фактчекинг». Мнения исследователей по данным терминам также разнятся. По этой причине каждый исследователь должен четко понимать, что он подразумевает под данными понятиями.

После определения аксиом исследователи журналисты непосредственно переходят к методам верификации информации. Большинство предлагаемых методов, относятся к одному из двух направлений: верификация источника и верификация контента.

Так в «Пособии по верификации информации» авторы рекомендуют традиционные принципы верификации, такие как «скептическая настроенность журналиста к получаемой информации», работа над «поиском достоверных источников», развитие «сети надежных источников», «постоянный поиск новых инструментов верификации, общение с коллегами».

Также в данной главе рассмотрены методы предлагаемые сферой машинного обучения для решения данной задачи.

Данную задачу можно отнести к задачам классификации или кластеризации. В первую очередь эти задачи отличаются подходом обучения моделей, а именно «с учителем и «без учителя». В случае обучения с «учителем» в модель подаются правильные ответы, по которым модель вычисляет ошибку и корректирует свои веса. В случае обучения «без учителя», модель сама определяет важные признаки, агрегирующие входные данные. Одним из основополагающих отличий при обучении «без учителя» является требование большого объема данных, что порой не всегда возможно. Именно по этой причине было решено решать задачу классификации новостного контента.

Одним из наиболее результативных инструментов машинного обучения являются глубокие нейронные сети. Одной из особенностей нейронных сетей является то, что они не умеют улавливать смысл текстовой информации в чистом виде. Именно по этой причине были разработаны методы, преобразующие текстовую информацию в числовую. Примерами таких методов являются «вложение слов», TF-IDF, «мешок слов» и др.

Также часто используемым методом передачи смысловой нагрузки текста является лемматизация. Данных подход приводит слова к начальной форме тем самым, упрощая процедуру преобразования текстовой информации: количество уникальных слов уменьшается, при этом смысл слов сохраняется.

Вторая глава посвящена более подробному анализу существующих решений классификации новостного контента и текстовой информации в частности.

Также более подробно рассматриваются инструменты глубокого обучения, а именно слои нейронных сетей, которые используются для обработки текстовой информации. Такими слоями являются «рекуррентные» и «сверточные» слои.

Особенностью рекуррентных слоев является их память о предыдущих результатах, что довольно важно при обработке информации, в которой важен порядок. Примерами рекуррентных слоев могут выступать «LSTM», «GRU» слои и их модификации.

Особенностью сверточных нейронных сетей является их подход в обработке информации, а именно уменьшение и увеличение размерности данных. Также сверточные сети часто состоят из блоков, в которые обычно входят свертночный слой, пулинг слой (MaxPooling, AveragePooling) и слой «dropout».

В данной главе представлены примеры применения данных методик при классификации текстовой информации.

В **третьей главе** предложены методы и алгоритмы реализации системы обучения моделей, а также пример реализации системы предсказания подлинности новостного контента в виде веб-приложения.

Система обучения моделей поделена на этапы и описывает полный процесс по подготовке данных, созданию и обучению моделей. Выводом системы является обученная модель, готовая к интегрированию в веб-приложение.

Главной особенностью предлагаемых моделей классификации новости является наличие трех «входов»: заголовок статьи, текст статьи, а также рейтинг авторов статьи. Такая модель помогает проанализировать каждую составляющую статьи по отдельности, после чего проверить их взаимосвязь – это особенно это касается заголовка и текста статьи. Ввиду того, что рейтинг авторов статьи представлен одним числом и, вообще говоря, является необязательным параметром (точнее, если авторы не указаны, то берется значение по умолчанию), ход исследования был сосредоточен на обработке текстовой информации.

Отдельно стоит выделить наличие 5 классов статей на выходе: 0 — новость фальшивая, 1 — скорее всего фальшивая новость, 2 — сложно определить достоверность или недостоверность новости, либо новость верна на половину, 3 — скорее всего новость правдива, 4 — новость правдива.

Ввиду большого количества возможных вариантов построения моделей были разработаны алгоритмы построения моделей, а также разработаны общие схемы построения моделей проводимых исследований. В качестве примера общая схема рекуррентных моделей представлена на рисунке 1.

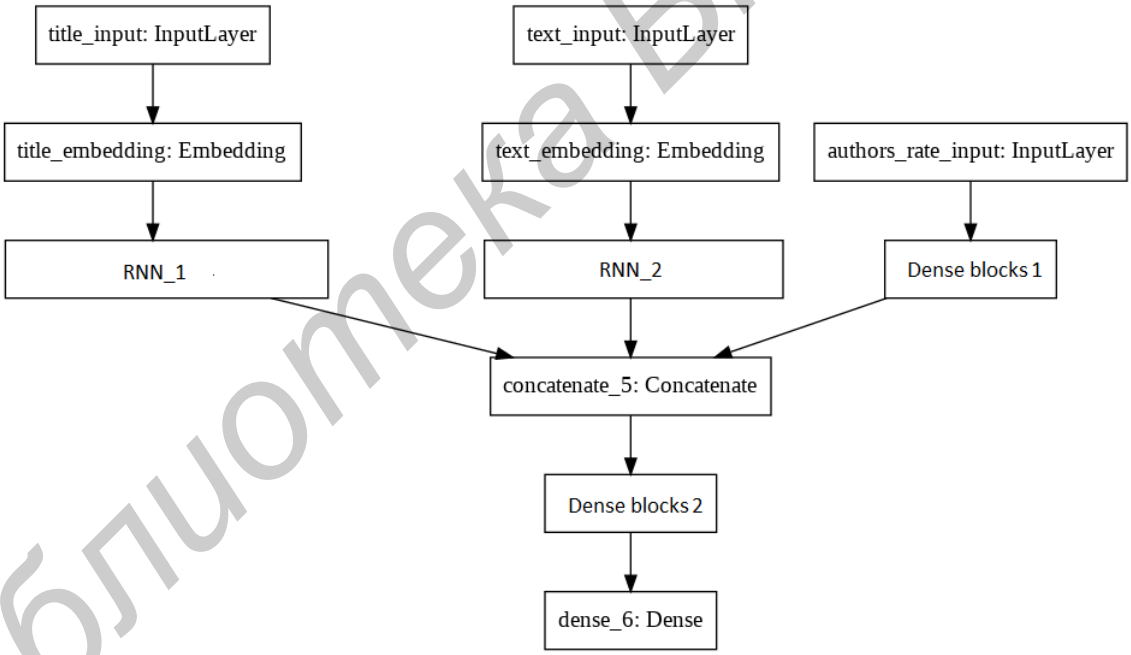

Рисунок 1 – Общая схема рекуррентных моделей

Реализованное веб-приложение состоит из удобного АПИ, к которому можно обращаться напрямую, а также из визуальной части. Реализованное АПИ содержит 2 эндпоинта: первый отвечает за классификацию новостного контента, а второй проверяет статус запроса на классификацию. Процесс классификации происходит в бэкграунде.

В **четвертой главе** описан ход исследования моделей, а также представлены и проанализированы результаты исследования.

Стоит отметить, что, ввиду большого количества всевозможных вариантов построения моделей, были исследованы наиболее часто используемые структуры нейронных сетей при решения задач, связанных с текстовой

информацией, а именно с применением рекуррентных и сверточных слоев. Также стоит отметить, что были исследованы и проанализированы модели, основанные на применении разных методов преобразования текстовой информации, а именно Word2Vec, GloVe и TF-IDF.

Наилучшие результаты показали модели «1blstm_8_8» и «cnn_256-1024_128-1024_mp_2» в своих категориях, при этом рекуррентная модель показала немногим лучше результат, нежели модель, основанная на сверточных слоях при использовании методов Word2Vec и GloVe, однако при использовании метода TF-IDF результат оказался обратным. Наилучшая точность модели «cnn_256-1024_128-1024_mp_2» составила 92.4%, а «1blstm_8_8» модели 92.1%.

Отдельно стоит отметить результативность применения метода TF-IDF относительно остальных методов преобразования текстовой информации. Результаты улучшились на 5%.

Также в данной главе была описана и проиллюстрирована логика работы веб-приложения, которое отвечает за прогнозирование подлинности новостного контента.

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

1. Исследованы методы решения задачи обнаружения фальшивых новостей. В частности, рассмотрены возможности теоретического анализа новостного контента, а также методы предлагаемые сферой машинного обучения, а именно нейронные сети. Рассмотрены некоторые модификации моделей, основанные на рекуррентных нейронных слоях, а также на сверточных нейронных слоях. Исследованы и проанализированы результаты обучения моделей, основанных на применении разных методов преобразования текстовой информации в числовую, а именно «вложение слов» с предобученными матрицами «Word2Vec» и «GloVe», а также TF-IDF метод. Наилучшие результаты показала модель «cnn_256-1024_128-1024_mp_2» с использованием TF-IDF в качестве метода преобразования текстовой информации.
2. Исследованы перспективы и причины для применения технологий машинного обучения в решении вопроса фальшивых новостей и их обнаружения в частности.
3. Разработана и обоснована архитектура программного средства, призванного решить проблему обнаружения фальшивых новостей. Данная система реализована в качестве веб-приложения.

Рекомендации по практическому использованию результатов

1. Полученные результаты формируют достаточную теоретическую и практическую базу для разработки программного обеспечения на основе технологий машинного обучения с применением нейронных сетей в частности.

Кроме того, они могут быть использованы для дальнейшего исследования и развития представленной системы.

2. . Разработанная архитектура приложения и подходы могут применяться не только в поставленной задаче, но и в любой другой, направленной на обработку текстовых данных и прогнозирование исследуемого явления. На основе разработанной архитектуры возможно создание системы любой сложности и масштаба.

3. Результаты работы могут использоваться при подготовке специалистов в области программного обеспечения и машинного обучения.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1-А. Губский М.Д. Применение нейронных сетей в решении задачи обнаружения фальшивых новостей // Интернаука: электрон. научн. журн. 2020. № 20(149). – Режим доступа: http://internauka.org/journal/science/internauka/149.