

ИССЛЕДОВАНИЕ ПРИМЕНИМОСТИ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ РАСПОЗНАВАНИЯ НЕДОСТОВЕРНОЙ ИНФОРМАЦИИ

INVESTIGATION OF THE ARTIFICIAL INTELLIGENCE TECHNOLOGIES APPLICABILITY FOR RECOGNIZING WRONG INFORMATION

V. Pavlov
V. Simdiankin

Summary. The paper studies the applicability of artificial intelligence and machine learning technologies for the analysis of publications of various information resources for recognizing inaccurate information on the Internet. It is shown that an artificial neural network can both be actively used as a tool to prevent a possible disinformation from various media, and be a part of a hardware-software complex for autonomous monitoring of Internet publications without the direct participation of an operator.

Keywords: neural network, machine learning, Internet, fake news, disinformation.

Павлов Виталий Юрьевич

К.ф.-м.н., доцент, Московский авиационный институт (национальный исследовательский университет)
vitaly_pavlov@hotmail.ru

Симдянкин Владимир Дмитриевич

Московский авиационный институт (национальный исследовательский университет)
leesadik@gmail.com

Аннотация. В работе исследована применимость технологий искусственного Интеллекта и машинного обучения для анализа публикаций различных информационных ресурсов для распознавания недостоверной информации в сети интернет. Показано, что искусственная нейронная сеть может как активно применяться в качестве инструмента предотвращения возможного факта дезинформации со стороны различных средств массовой информации, так и входить в состав аппаратно-программного комплекса по автономному мониторингу интернет-изданий без непосредственного участия оператора.

Ключевые слова: нейросеть, машинное обучение, Интернет, ложные новости, дезинформация.

Искусственные нейронные сети находят всё более широкое применение для решения задач, ранее доступных только человеку. Оценка достоверности информации и распознавание дезинформации является одной из таких задач, поскольку искусственная нейронная сеть является математической моделью, построенной по принципу организации и функционирования биологических нейронных сетей — систем нервных клеток живого организма, в том числе, человеческого мозга. Создание в последние годы высокопроизводительных вычислительных комплексов в совокупности с усовершенствованием алгоритмов обучения позволили использовать нейросетевые модели в практических целях: в задачах прогнозирования, для распознавания образов, в задачах управления, автоматического перевода и др.

Искусственная нейронная сеть (ИНС) представляет собой систему соединённых и взаимодействующих между собой простых процессоров (искусственных нейронов). Такие процессоры обычно довольно просты (особенно в сравнении с процессорами, используемыми в персональных компьютерах). Каждый процессор подобной сети имеет дело только с сигналами, которые

он периодически получает, и сигналами, которые он периодически посылает другим процессорам. И, тем не менее, будучи соединёнными в достаточно большую сеть с управляемым взаимодействием, такие по отдельности простые процессоры вместе способны выполнять довольно сложные задачи.

Нейронные сети используются для решения сложных задач, которые требуют аналитических вычислений подобных тем, что делает человеческий мозг. Самыми распространёнными применениями на практике нейронных сетей является:

Классификация — распределение каких-либо данных по параметрам. Например, на вход дается набор людей и нужно решить, кому из них давать кредит, а кому нет. Эту работу может сделать нейронная сеть, анализируя такую информацию как: возраст, платежеспособность, кредитная история и т.д.

Предсказание — возможность предсказывать следующий шаг. Например, рост или падение акций, основываясь на ситуации на фондовом рынке.

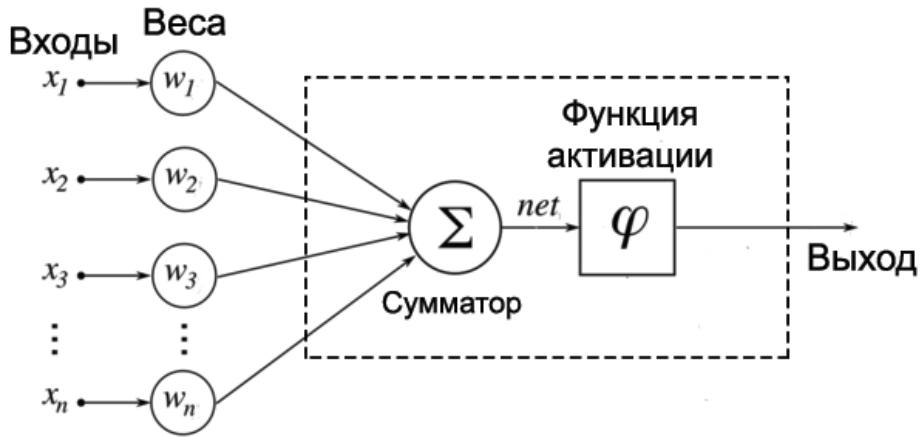


Рис. 1. Схематическая модель нейрона

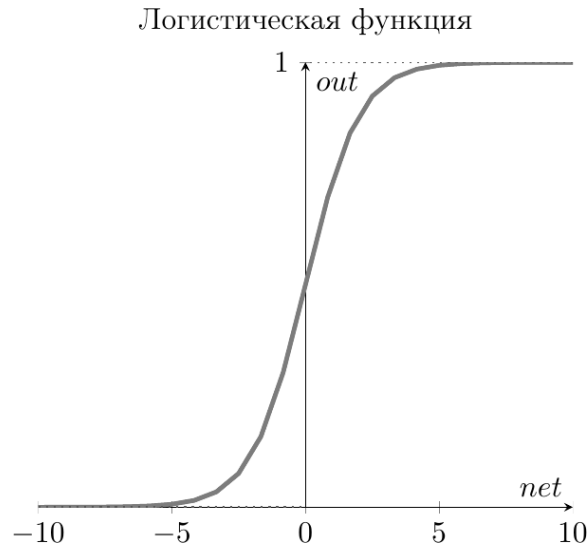


Рис. 2. График логистической (сигмоидальной функции)

Распознавание — самое широкое применение нейронных сетей, в настоящее время. Используется в Google, когда вы ищете фото или в камерах телефонов, когда оно определяет положение вашего лица и выделяет его и многое другое.

В данной работе нас непосредственно интересует функция классификации. Во время исследования по данной теме самым важным будет определение параметров, по которым в дальнейшем будет производиться сама классификация новостей. В качестве параметров можно брать авторитет издательства(переменная в промежутке от 0 до 1, которая в свою очередь показывает насколько можно доверять публикациям от тех или иных издательств, 0 — практически всегда издательство публикует недостоверную информацию, 1 — практически все публикации являются правдивыми), так же мож-

но использовать базу данных в которой будут собраны все авторы с низким рейтингом доверия.

Нейрон — это вычислительная единица, которая получает информацию, производит над ней простые вычисления и передает ее дальше.

Если мы будем просто подавать какую-либо взвешенную сумму на выход, то какого-либо осмысленного результата мы не получим. Нейрон должен как-то обработать ее и сформировать адекватный выходной сигнал. Именно для этих целей обычно используют функцию активации.

Она преобразует взвешенную сумму в какое-то число, которое и является выходом нейрона (выход нейрона обозначим переменной *out*)

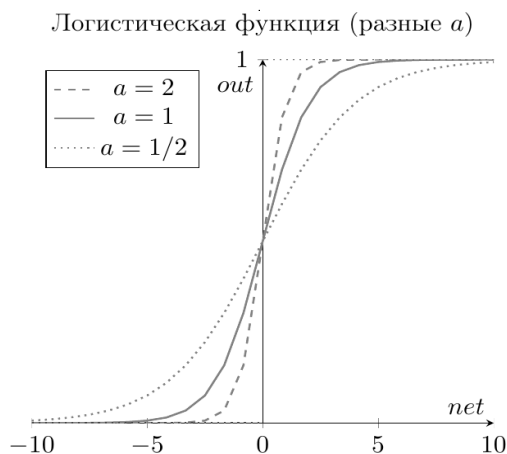


Рис. 3. График логистической функции с различными параметрами a .

Для разных типов искусственных нейронов используют самые разные функции активации. В общем случае их обозначают символом $\phi(net)$. Указание взвешенного сигнала в скобках означает, что функция активации принимает взвешенную сумму как параметр.

Существует три наиболее часто используемые функции активации

- ◆ Функция единичного скачка
- ◆ Сигмоидальная функция
- ◆ Гиперболический тангенс

Конкретно в нашем случае наиболее подходящей является именно сигмоидальная функция. Давайте рассмотрим ее поподробнее.

График этой функции выглядит достаточно просто. А вот так она записывается аналитически:

$$out(net) = \frac{1}{1 + \exp(-a \cdot net)}$$

Параметр a это какое-то число, которое характеризует степень крутизны функции. Ниже представлены логистические функции с разным параметром a .

Вспомним наш искусственный нейрон, определяющий, является ли новость достоверной или нет. В случае с функцией единичного скачка все было очевидно. Новость либо недостоверная (1), либо достоверная (0).

Здесь же случай, более приближенный к реальности. Мы до конца полностью не уверены, достоверная наша новость или нет. Тогда использование логистической функции в качестве функции активации приведет к тому, что вы будете получать цифру между 0 и 1. Причем чем больше взвешенная сумма, тем ближе выход будет к 1

(но никогда не будет точно ей равен). И наоборот, чем меньше взвешенная сумма, тем ближе выход нейрона будет к 0.

Например, выход нашего нейрона равен 0.8. Это значит, что он считает, что новость практически точно недостоверная. Если бы его выход был бы равен 0.2, то это означает, что новость почти наверняка достоверная.

Какие же свойства имеет логистическая функция?

- ◆ она является «сжимающей» функцией, то есть вне зависимости от аргумента (взвешенной суммы), выходной сигнал всегда будет в пределах от 0 до 1
- ◆ она более гибкая, чем функция единичного скачка — ее результатом может быть не только 0 и 1, но и любое число между ними
- ◆ во всех точках она имеет производную, и эта производная может быть выражена через эту же функцию

Именно из-за этих свойств логистическая функция чаще всего используется в качестве функции активации в искусственных нейронах.

В качестве модели ИНС разумно использовать модель перцептрона.

Перцептрон (Perceptron) — простейший вид нейронных сетей. В основе лежит математическая модель восприятия информации мозгом, состоящая из сенсоров, ассоциативных и реагирующих элементов.

Под многослойным перцептроном понимают два разных вида: многослойный перцептрон по Розенблатту и многослойный перцептрон по Румельхарту. Многослойный перцептрон по Розенблатту содержит более 1 слоя А-элементов. Многослойный перцептрон по Ру-

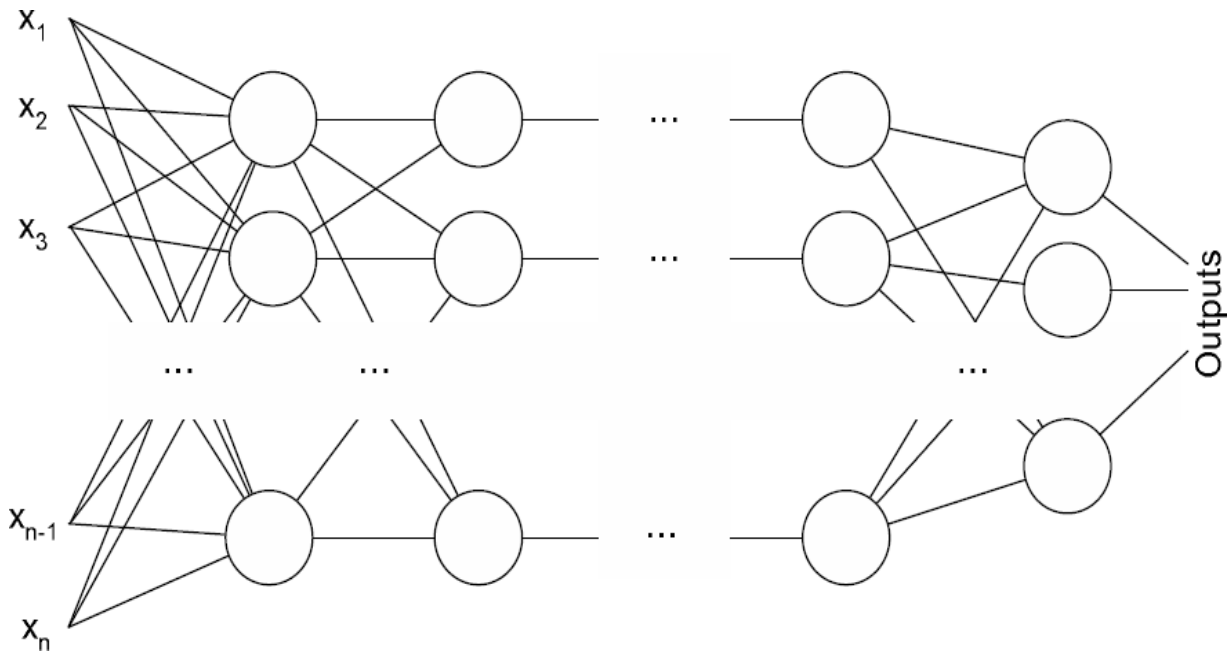


Рис. 4. Алгоритм работы многослойного персептрона.

мельхарту является частным случаем многослойного персептрона по Розенблатту, с двумя особенностями:

- ◆ S-A связи могут иметь произвольные веса и обучаться наравне с A-R связями.
- ◆ Обучение производится по специальному алгоритму, который называется обучением по методу обратного распространения ошибки.

Этот метод является краеугольным камнем обучения всех многослойных ИНС. Во многом благодаря ему возобновился интерес к нейронным сетям.

Самая популярная и достаточно устаревшая архитектура, в которой подряд идут несколько слоев нейронов — входной, один или несколько скрытых слоев, и выходной слой.

Практически всегда обучается методом обратного распространения ошибки — что означает, что мы должны предоставить для обучения сети набор пар «входной вектор — правильный выход». В таком случае входной вектор будет отправлен на вход сети, последовательно будут рассчитаны состояния всех промежуточных нейронов, и на выходе образуется выходной вектор, который мы будем сравнивать с правильным. Разница даст нам ошибку, которую в дальнейшем можно будет распространить в обратном порядке по связям сети, высчитать вклад в итоговую ошибку каждого нейрона, и скорректировать его веса, чтобы ее исправить. Если повторить данную процедуру n-ое количество раз, то у нас получится обучить сеть. Количество дополнительных слоев будет влиять на точность работы нашей сети.

Сеть такого типа обычно подходит для ряда следующих случаев и в свою очередь хорошо справляется с задачами, где:

1. Если ответ действительно зависит только от того, какие входные данные получает сеть, и никак не зависит от истории входных данных, которые сеть получала ранее (т.е. это не динамический процесс, или, по крайней мере, мы дали на вход исчерпывающую информацию об этом процессе в форме, пригодной для обработки сетью).
2. Ответ не зависит/слабо зависит от высоких степеней и/или произведений параметров — функции этого типа сеть строить почти не умеет.
3. В наличии имеется достаточное количество примеров (желательно иметь не менее сотни примеров на каждую связь сети), или у вас имеется достаточный опыт борьбы с эффектом специализации. Это в большинстве случаев связано с тем, что имея достаточно много коэффициентов, сеть может банально только запомнить достаточное количество конкретных примеров, и выдавать на них более точный результат — но прогнозы нашей сети не будут иметь совершенно ничего общего с реальностью в случае, если мы попытаемся дать на вход примеры, которые отсутствуют в обучающей выборки.

Данный класс нейронных сетей достаточно изучен со всех сторон и известны все аспекты его работы, хорошо отрабатывает на своих задачах, если на некоторой задаче не работает — то это причина для того чтобы сделать вывод, что данная задача которую пытаются решить

по средствам нейронной сети относящийся к данному классу по факту оказалась сложнее, чем казалось.

Но также, данный класс нейронных сетей имеет слабые стороны, а именно это неумение работать с динамическими процессами, и потребность в большой обучающей выборке для более точного результата. Таким образом выглядят технические требования по устройству ИНС.

Рассмотрим возможные признаки для выявления недостоверных новостей.

1. Гиперболизированные и кликбейтные заголовки.

“Трамп в шоке от нового закона. ВЫ ЧТО СДУРЕЛИ!!?”

Новостные заголовки обычно передают суть информационного сообщения, а не создают загадок

“УМЕР СОЛИСТ ГРУППЫ СПЛИН. Вся страна в глубоком отчаянии.”

Когда умирает действительно «легендарный рок-музыкант, хиты которого знают все» достаточно написать его имя, чтобы привлечь внимание читателей

Заголовки новостей коротко описывают суть произошедшего. В противном случае лучше новость проигнорировать.

2. Подозрительные домены или искажённые названия сайтов популярных медиа, похожие на настоящие

Забейте в поисковик название интересующего вас СМИ — на первых позициях в выдаче будут оригиналы. Сравните их интернет-адресом подозрительного источника.

3. Большое количество ошибок или опечаток в тексте новости

Попробуйте найти на сайтах известных вам СМИ эту же историю, но в более развёрнутом виде. Если поиски не увенчаются успехом, то, скорее всего, перед вами автоматически сгенерированный контент.

4. Поддельные фото или коллажи

Фотоколлаж рассчитан на невнимательного читателя и призван иллюстрировать несостоявшееся событие

Попробуйте сохранить подозрительное фото и поискать его копии через фото поисковики (Google, TinEye

и т.п.). Если копий нет в авторитетных источниках, то скорее всего перед вами фотошоп.

5. Отсутствие отметок о времени публикации новости

Даты публикации скрывается для того, чтобы максимально продлить «жизнь» фейковой новости. Такое сообщение будет казаться актуальным сколь угодно долго, пока не будет явно противоречить действительности.

Забейте заголовок подозрительной новости и попробуйте поискать его, используя фильтры времени. Скажем, укажите прошлый год или прошлый месяц. Если вы найдёте аналогичную новость в архивах, то скорее всего это фейк.

6. Отсутствие указания автора и источника

Многие фейки генерируются автоматизировано, поэтому никаких упоминаний ни автора, ни источника не остаётся. Тем не менее в последнее время фейковые новости выходят за авторством никому не известных журналистов — скорее всего их имена также случайны, и за ними не стоят реальные люди.

Попытайтесь найти новость или её авторов в мейнстрим-медиа. Прежде всего — в информационных агентствах.

Если же источники указаны, то попробуйте изучить их. Источники типа «Wikipedia» или «LiveJournal» — это скорее попытки скрыть реальные источники информации или их отсутствие.

7. Расставление весов по каждому из критериев (в диапазоне от 0 до 1)

8. Построение вектора и передача его в функцию активации

9. Результат работы функции активации (переменная со значением из диапазона от 0 до 1) будет показывать, насколько новость является достоверной.

В результате анализа применимости ИНС для распознавания недостоверных и ложных новостей можно сделать следующие выводы.

1. Несмотря на отсутствие в настоящий момент однозначного определения понятия «фейк-ньюз», «фальшивые новости» набирают обороты и требуют дальнейшего изучения [1–3].
2. Медиаорганизации и технологические компании пытаются определить, как лучше всего ре-

агировать на фейковые новости. Возможно, помогут более явная проверка фактов, улучшение медиаграмотности населения, дополнительные инвестиции государства в образование, медиа, расследовательский журнализм и различные социальные институты, с целью обучить граждан правильно анализировать прессу, чтобы противостоять дезинформации [4–5].

3. Исследователи соглашаются с необходимостью разработки автоматизированных вычислительных систем для идентификации поддельной информации в социальных сетях [6–7]. Наше исследование демонстрирует одно из возможных решений проблемы за счет методов определения

этого типа новостей в социальных сетях, основанных на использовании систем с элементами искусственного интеллекта и машинного обучения. Применение искусственного интеллекта в связке с инструментарием по обработке веб-данных имеет большой потенциал [8–9].

4. Мы должны быть крайне осторожны в своих выводах, потому что они очень чувствительны к выборке ложных и реальных источников новостей. Сбор и анализ эмпирических данных о фейковых новостях, по сути, только начинается. Наше понимание этого нового явления по-прежнему весьма ограничено, но, несомненно, будет улучшаться в будущем

ЛИТЕРАТУРА

1. Fake news and disinformation online. Report. European Union, 2018. URL: <https://ec.europa.eu/digital-single-market/en/news/final-resultseurobarometer-fake-news-and-online-disinformation>
2. Bentzen N. Understanding disinformation and fake news. European Parliamentary Research Service. 2017. URL: [http://www.europarl.europa.eu/RegData/etudes/ATAG/2017/599408/EPRS_ATA\(2017\)599408_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/ATAG/2017/599408/EPRS_ATA(2017)599408_EN.pdf)
3. Jowett G., O'Donnell V. Propaganda and Persuasion. SAGE Publications, 2015. 480 p.
4. Greenberg A. Inside Google's Internet Justice League And Its AiPowered War On Trolls // Wired. 19.09.2016. URL: <https://www.wired.com/2016/09/inside-googles-internet-justiceleague-ai-powered-war-trolls/>
5. Filatova O., Bolgov R. Strategic Communication in the Context of Modern Information Confrontation: EU and NATO vs Russia and ISIS // Proceedings of the 13th International Conference on Cyber Warfare and Security ICCWS2018. Edited by Dr John S. Hurley and Dr Jim Q. Chen. Washington DC, USA, 2018. P. 208–219.
6. Shu K., Sliva A., Wang S., Tang J., Liu H. Fake News Detection on Social Media: A Data Mining Perspective // Arxiv.org. URL: <https://arxiv.org/pdf/1708.01967.pdf>
7. Koshy J. How Content Discovery Platforms Can Fight Fake News via Web Scraping and AI // PromtCloud. URL: <https://www.promptcloud.com/blog/fight-fake-news-web-scrapingartificial-intelligence>
8. Некрасов Г.А., Романова И.И. Разработка поискового робота для обнаружения веб-контента с фейковыми новостями // Инновационные, информационные и коммуникационные технологии. 2017. № 1. С. 128–130.
9. Жук Д.А., Жук Д.В., Третьяков А.О. Методы определения поддельных новостей в социальных сетях с использованием машинного обучения // Информационные ресурсы России. 2018. № 3. С. 29–32.

© Павлов Виталий Юрьевич (vitaly_pavlov@hotmail.ru), Симдянкин Владимир Дмитриевич (leesadik@gmail.com).

Журнал «Современная наука: актуальные проблемы теории и практики»