

бесконечном множестве научной литературы, подборе наиболее актуальных и авторитетных источников информации.

Заключение. Несмотря на возможности отслеживания тенденций практически по всем научным направлениям, созданные механизмы представляют наибольший интерес для авторов в сфере цифрового развития, поскольку подобные решения позволяют на ранних этапах отслеживать формирование новых технологий, а также обеспечить быстрое реагирование на их изменения.

ЛИТЕРАТУРА

1. Севрук Е.А., Качан Д.А., Юневич Н.Г. Выделение трендов цифрового развития с помощью автоматизированной аналитики частоты и тематики научных публикаций // Сборник докладов XXII Международной конференции «РИНТИ-2023». – 2023. – С. 148–153.

2. ResearchGate: Find and share research [Electronic resource]. – URL: <https://www.researchgate.net/>. – Дата доступа: 03.01.2024.

3. Interactive network visualizations. Pyvis Documentation [Electronic resource]. – URL: <https://pyvis.readthedocs.io/en/latest/> – Дата доступа: 03.01.2024.

4. Cyberleninka.ru [Electronic resource] // Similarweb: Website Traffic – Check and Analyze Any Website: [сайт]. – URL: <http://east-front.narod.ru/memo/latchford.htm> – Дата доступа: 03.01.2024.

5. Axios: [сайт]. – URL: <https://axios-http.com/docs/intro>.

6. Cheerio.js: [сайт]. – URL: <https://cheerio.js.org/>.

УДК 004. 032.26

Д.В. Шиман, доц., канд. техн. наук;
А.Н. Мушук, ассист.; Е.А. Гончар, ассист.
(БГТУ, г. Минск)

ЭФФЕКТИВНОСТЬ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ СОЦИАЛЬНЫХ ЯВЛЕНИЙ

На данный момент существует большое количество различных алгоритмов машинного обучения. Каждый из них был создан под разные задачи. В свою очередь, задачи делятся на подвиды. Одним из которых является прогнозирование.

В целом, практически любой алгоритм можно использовать для решения любого типа задач. Отличия состоят лишь в том, что точность работы алгоритма будет слегка ниже, чем если бы использовался предназначенный для конкретной задачи, однако иногда этим мож-

но пренебречь. Однако для достижения наилучшего результата, само собой, требуется, как минимум, использовать готовый, специализирующийся на решении конкретного типа задач, алгоритм. Еще лучшего результата можно добиться, если не просто взять готовый, а создать алгоритм под конкретную задачу.

В случае нейронных сетей, под созданием алгоритма обычно подразумевается выбор слоев, количества нейронов в каждом, функций активации и значение обновления весов.

Стояла задача, разработать алгоритм, способный спрогнозировать количество самоубийств, по социально-экономическим показателям стран. Предложенный алгоритм имеет два слоя длинной краткосрочной памяти, в каждом по восемьдесят нейронов. И два полносвязных слоя, один из которых содержит двадцать нейронов и служит для промежуточного объединения полученных данных, а второй содержит один нейрон и необходим исключительно для вывода информации. Полученная архитектура представлена на рис. 1.

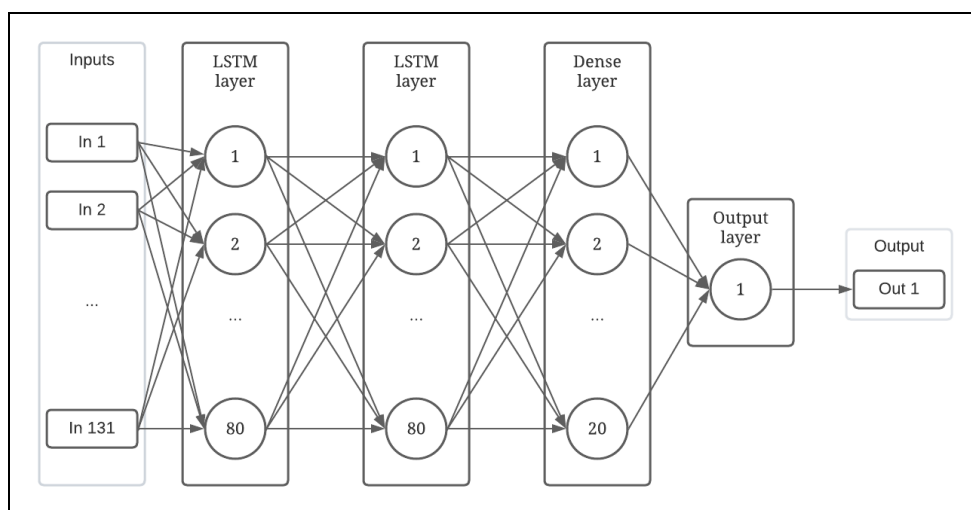


Рисунок 1 – Архитектура сети

Точность прогноза данной сети можно оценить как достаточно высокую. Так, среднее значение показателя средней абсолютной ошибки для десяти циклов обучения на случайно перемешанных данных составило 2.6629. В то время, как алгоритм случайного леса, с количеством деревьев равным ста, на тех же данных показал значение 2.8499.

Прогнозирование социальных явлений, в целом, является не самой распространенной темой исследований. Их количество в этой области достаточно невелико. Тем не менее, они существуют. Однако следует учитывать контекст. Рассматриваемые далее алгоритмы настраивались каждый под свою задачу. Их все еще можно сравнивать с предложенным, так как область и задача весьма схожи, однако не

идентичны. Соответственно в каждом будут приведены значения, полученные в рассматриваемых исследованиях с данными исследований. Сравнение же будет проводиться как с результатами исследований, так и с результатами, полученными при применении полностью идентичных алгоритмов к нашему набору данных, где это возможно.

Первым исследованием, с которым стоит сравнить эффективность полученной сети, является исследование, посвященное прогнозированию попыток самоубийства среди подростков [1]. В данном исследовании использовались четыре алгоритма: логистическая регрессия, алгоритмы Lasso и Ridge, и случайный лес. Показателем эффективности авторы выбрали AUC (Area Under Curve). Этот показатель считает площадь под кривой результатов. И, упрощая, чем он выше, тем более точный результат был получен. Для четырех, указанных выше, алгоритмов показатель AUC составил 0.828, 0.826, 0.829, и 0.824 соответственно.

Следующим рассматриваемым исследованием является сравнение моделей прогнозирования риска суицидального поведения у корейских подростков [2]. В данной работе были рассмотрены алгоритмы логистической регрессии, случайного леса, машины опорных векторов, искусственной нейронной сети и XGB (eXtreme Gradient Boosting). Рассматриваемая авторами нейронная сеть является однослойной, обучаемой на шестидесяти эпохах и оптимизируется алгоритмом Adam. Для оценки эффективности в данном исследовании также взят показатель AUC, и он составил 0.851, 0.853, 0.857, 0.851 и 0.863 соответственно. Применимо к набору данных, предназначенного для предложенного алгоритма, AUC логистической регрессии, случайного леса и нейронной сети составил 0.4944, 0.7021 и 0.8311 соответственно.

Еще одним рассматриваемым исследованием является прогноз смертности от самоубийств на уровне штата в США [3]. Оно представляет наибольший интерес, так как в данном исследовании рассматривается эффективность нейронной сети схожей по архитектуре с предложенной. Авторы проводят сравнение нескольких конфигураций сети, и наилучшие результаты были получены на сети с одним LSTM слоем, количеством нейронов равным 32, 16, 64, 16 для Юты, Луизианы, Нью-Йорка и Колорадо и количеством эпох обучения 200, 200, 250, 150 для тех же штатов. Функции активации и повторной активации во всех случаях представлены тангенсом и сигмоидой соответственно. Метрикой точности сети в данном исследовании был корень среднеквадратичной ошибки (RMSE – Root Mean Square Error). Сравнение результативности представлено в табл. 1.

Таблица 1 – Сравнение результативности

Количество нейронов	Количество эпох обучения	RMSE на предложенных авторами данных	RMSE на предложенных данных
32	200	3.765	6.9755
16	200	4.156	6.5639
64	250	7.414	6.6872
16	150	5.889	6.3545

Можно заметить, что результативность предложенных авторами исследований алгоритмов заметно ниже, при применении их к нашему набору данных. Однако в случае применения предложенного алгоритма, обрабатывающего наш набор данных, показатель AUC равен 0.8804, что примерно на 0.05 выше, чем аналогичные показатели в первом исследовании и на 0.03 выше, чем во втором.

Показатель RMSE равен 5.4683, что чуть хуже двух из четырех алгоритмов предложенных авторами третьего исследования. Однако, как уже упоминалось, каждый алгоритм создается под конкретную задачу и набор данных, и алгоритмы из третьего исследования дают результаты значительно хуже, чем предложенный, на нашем наборе данных. Что может говорить о его эффективности в контексте поставленной задачи.

Подводя итог, предложенный алгоритм предсказывает значения в целом практически всегда лучше, чем аналоги, а на своем наборе данных значительно лучше. Однако стоит отметить, что прогнозирование социальных явлений – это сложная и многогранная задача, требующая для решения не только разработки нейронной сети, но и полного понимания социальных, психологических и экономических факторов, влияющих на это явление. Нейронные сети могут быть полезным инструментом, но необходимо учитывать ограничения и контекст, в котором они применяются.

ЛИТЕРАТУРА

1 Miche M. et al. Prospective prediction of suicide attempts in community adolescents and young adults, using regression methods and machine learning //Journal of affective disorders. – 2020. – Т. 265. – С. 570–578. – DOI: doi.org/10.1016/j.jad.2019.11.093.

2 Jung J.S. et al. Prediction models for high risk of suicide in Korean adolescents using machine learning techniques // PLoS one. – 2019. – Т. 14. – №. 6. – С. e0217639. – DOI: doi.org/10.1371/journal.pone.0217639.

3 Patel D. et al. Predicting state level suicide fatalities in the united states with realtime data and machine learning //npj Mental Health Research. – 2024. – Т 3. – №. 1. – С. 3. – DOI: doi.org/10.1038/s44184-023-00045-8