

Подход к эффективной обработке информации с датчиков системы умного дома при помощи методов машинного обучения

И. Б. Саенко¹, О. С. Лаута², Н. А. Васильев³, В. Е. Садовников³

¹ Санкт-Петербургский Федеральный исследовательский центр Российской академии наук

² Государственный университет морского и речного флота имени Адмирала С.О. Макарова

³ Военная академия связи имени Маршала Советского Союза С.М. Буденного

ibsaen@comsec.spb.ru

Аннотация. В статье проведен анализ различных алгоритмов машинного обучения, которые могут применяться для обработки информации с датчиков системы умного дома. На основе результатов проведенного анализа предложен новый подход к эффективной обработке информации, получаемой с датчиков умного дома, основанный на использовании архитектуры сверточной нейронной сети MobileNet и программной библиотеки для машинного обучения TensorFlow, а также средства мониторинга TensorBoard. Эксперименты показали, что предлагаемый подход позволяет достичь достаточно высокой точности и эффективности обработки информации.

Ключевые слова: машинное обучение; умный дом; сверточная нейронная сеть; обработка информации

I. ВВЕДЕНИЕ

В настоящее время системы умного дома приобретают все большую популярность, предоставляя уникальные возможности по автоматизации и контролю различных аспектов жизни в доме. Датчики, установленные в таких системах, играют ключевую роль в сборе информации о состоянии окружающей среды, энергопотреблении и поведении пользователей. Однако, с ростом количества собираемых данных возникает необходимость эффективной обработки и анализа этой информации для принятия обоснованных и рациональных решений [1].

В последние годы методы машинного обучения получили широкое применение в области анализа данных, позволяя автоматически извлекать полезную информацию из больших объемов неструктурированных данных. Использование методов машинного обучения в обработке информации с датчиков системы умного дома открывает новые возможности для оптимизации работы и повышения эффективности таких систем. Однако, в существующих исследованиях в этой области недостаточно внимания уделяется разработке подходов, специализированных и адаптированных именно для обработки данных с датчиков умного дома [2].

Сегодня тематика умных домов становится все более актуальной благодаря развитию микроэлектроники, каналов связи, Интернет-технологий и Искусственного Интеллекта. Жилище человека претерпело значительные изменения на протяжении времени. В эпоху

Промышленной Революции 4.0 и Интернета Вещей оно стало комфортным, удобным и безопасным. На рынке появляются решения, которые превращают квартиру или загородный дом в сложные информационные системы, которыми можно управлять из любой точки мира.

В данной статье на основе результатов анализа известных алгоритмов машинного обучения предложен новый подход к эффективной обработке информации с датчиков системы умного дома, основанный на совместном использовании архитектуры сверточной нейронной сети MobileNet и программной библиотеки для машинного обучения TensorFlow, а также средства мониторинга TensorBoard. Экспериментальные результаты показали, что предлагаемый подход позволяет достичь достаточно высокой точности и эффективности обработки информации.

II. АНАЛИЗ ПОДХОДОВ К ПРИМЕНЕНИЮ МАШИННОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА ДАННЫХ УМНОГО ДОМА

Машинное обучение для анализа данных умного дома используется для обработки и анализа больших объемов информации, собираемой различными сенсорами и устройствами в доме.

Одним из примеров является анализ данных о потреблении энергии в доме. Модели машинного обучения могут использоваться для прогнозирования будущего потребления, определения временных пиков использования энергии, а также для оптимизации использования ресурсов. Например, они могут предложить оптимальное расписание работы устройств с учетом дешевых временных тарифов электроэнергии [3].

Другой пример – управление температурой и климатическими условиями в доме. С помощью машинного обучения можно создать модель, которая определит оптимальные настройки для системы отопления и охлаждения, исходя из различных факторов, таких как погода, время года, привычки проживающих в доме людей [3].

Кроме того, машинное обучение может быть применено для анализа данных о безопасности и защите дома. Например, модели машинного обучения могут анализировать видео с камер и определять подозрительную активность, обнаруживать вторжение или предупреждать о пожаре или утечке газа [3].

Важной частью машинного обучения для анализа данных умного дома является обучение моделей на

Исследование выполнено за счет гранта Российского научного фонда № 23-11-20024, <https://rscf.ru/project/23-11-20024/>, и Санкт-Петербургского научного фонда.

размеченных данных. Например, для надлежащего функционирования системы управления потреблением энергии и прогнозирования ее использования, необходимо иметь исторические данные о потреблении энергии в различных условиях [4].

Алгоритмы машинного обучения можно разделить на следующие классы:

- регрессионный анализ;
- кластеризация;
- классификация;
- нейронные сети;
- ассоциативные правила.

Алгоритмы регрессионного анализа используются для прогнозирования и предсказания значений переменных, таких как потребление энергии или температура в доме. Регрессионная модель может использоваться для определения зависимостей между входными переменными (например, погода, время суток) и выходными переменными (например, потребление энергии).

Кластеризация позволяет группировать различные устройства и сенсоры, а также пользователей, на основе их характеристик и поведения. Например, можно сгруппировать устройства, которые сопоставляются с определенными паттернами использования энергии, чтобы идентифицировать наиболее энергоэффективные устройства.

Классификация используется для разбиения определенных событий или состояний в умном доме на классы. Например, можно классифицировать активность людей в доме на "сон", "работа", "отдых" и т. д. Это поможет автоматически управлять системами умного дома в зависимости от текущего состояния [5].

Нейронные сети также можно использовать для анализа данных умного дома. Нейронные сети способны обрабатывать сложные и нелинейные данные. Например, сверточные нейронные сети можно использовать для обработки видео с камер и обнаружения подозрительной активности [6].

Ассоциативные правила используются для выявления различных ассоциаций и связей между устройствами и событиями в умном доме. Например, можно выяснить, что, когда свет включается в комнате, телевизор также часто включается в той же комнате, исходя из исторических данных.

Разберем более подробно методы каждого из подходов. Регрессионный анализ включает:

- линейную регрессию;
- логистическую регрессию;
- регрессию на деревьях решений;
- регрессию на нейронных сетях.

Линейная регрессия – это наиболее распространенный метод регрессионного анализа. Линейная регрессия моделирует линейную зависимость между зависимой переменной и одной или более независимых переменных. В системах умного дома линейная регрессия может быть использована, например,

для прогнозирования энергопотребления на основе погодных условий или количества жильцов в доме.

Логистическая регрессия используется для моделирования бинарной зависимой переменной. Логистическая регрессия может быть полезна в системах умного дома для предсказания наличия или отсутствия определенных событий или состояний, например, обнаружение движения или открытых дверей [7].

Регрессия с использованием деревьев решений позволяет разбить данные на структурированные подгруппы в виде дерева. Каждый узел представляет собой вопрос или условие, на которое отвечает наблюдение. Регрессия с использованием деревьев решений предсказывает континуальные значения зависимой переменной. В системах умного дома это может быть полезно для прогнозирования, например, энергоэффективности в зависимости от различных факторов, таких как типы устройств и их использование.

Нейронные сети, включая глубокие нейронные сети, могут использоваться для регрессионного анализа данных систем умного дома. Они обладают высокой способностью моделировать сложные зависимости между переменными и могут использоваться для прогнозирования энергопотребления или других характеристик системы [8].

Среди методов кластеризации выделяют следующие:

- k-средних;
- иерархическая кластеризация;
- DBSCAN;
- Гауссовы смешанные модели;
- спектральная кластеризация.

Наиболее распространенным является метод k-средних. Он разделяет данные на заранее определенное количество кластеров, где каждый кластер имеет центроид. В системах умного дома этот метод может использоваться для группировки похожих устройств, расположенных в разных комнатах или зонах [9].

Иерархическая кластеризация создает иерархическую структуру кластеров, начиная с отдельных объектов, объединяя их, чтобы образовать все более крупные кластеры. Этот подход может быть полезен для систем умного дома, чтобы классифицировать различные комнаты или устройства на основе их связи и близости друг к другу.

Метод DBSCAN (Density-Based Spatial Clustering of Applications with Noise) основан на плотности данных. Он определяет кластеры, исходя из плотности объектов в пространстве данных. DBSCAN может быть применен для кластеризации устройств с учетом схожести в их поведении и использовании [10].

Гауссовы смешанные модели (Gaussian Mixture Models, GMM) используют смесь нормальных распределений для представления данных. Они позволяют учитывать не только форму и структуру данных, но и взаимосвязь между кластерами. GMM могут быть использованы для кластеризации устройств, на основе их характеристик и поведения [11].

Метод спектральной кластеризации основана на спектральном анализе графа данных. Он строит граф, где

узлы представляют объекты, а ребра отражают их сходство. Затем осуществляется разделение графа на кластеры с использованием собственных векторов и значений. Этот метод может быть полезным для идентификации групп устройств или помещений на основе их взаимодействий.

Методы классификации включают:

- логистическую регрессию;
- метод опорных векторов;
- решающие деревья;
- гауссовы наивные Байесовские классификаторы;
- глубокие нейронные сети.

Хотя логистическая регрессия может использоваться для моделирования бинарной зависимой переменной, она может быть применена и для многоклассовой классификации. Логистическая регрессия может быть использована для классификации, например, типов устройств в системе умного дома, на основе их характеристик или действий.

В методе опорных векторов (Support Vector Machine, SVM) ищется оптимальная гиперплоскость, которая разделяет классы в пространстве данных. SVM может быть применен для классификации устройств или событий в системе умного дома, на основе их характеристик или поведения. [12]

Решающие деревья используются для создания дерева решений на основе различных условий. В системах умного дома решающие деревья могут быть использованы для классификации объектов, таких как устройства или состояния системы.

Гауссовы наивные Байесовские классификаторы основаны на предположении о независимости между признаками и на условных гауссовских распределениях. Они могут быть применимы для классификации устройств или состояний системы умного дома на основе их характеристик или событий.

Глубокие нейронные сети обладают способностью обрабатывать сложные зависимости и выявлять скрытые закономерности в данных. Они также могут быть использованы для классификации данных систем умного дома, что делает их мощным инструментом для классификации различных объектов и состояний в системе [13].

Нейронные сети включают в себя следующие методы:

- рекуррентные нейронные сети;
- сверточные нейронные сети;
- глубокие нейронные сети;
- автокодировщики;
- генеративно-состязательные сети.

Рекуррентные нейронные сети позволяют учитывать последовательность данных, что особенно полезно для анализа временных рядов в системах умного дома, таких как потребление энергии или температура. Эти сети могут обрабатывать входные данные с прошлых моментов времени и использовать эту информацию для прогнозирования будущего состояния системы [14].

Сверточные нейронные сети (Convolutional Neural networks, CNN) часто используются для анализа изображений, но также могут применяться к анализу данных сенсоров в системах умного дома. Они могут автоматически извлекать характеристики из сырых данных и использовать их для классификации или прогнозирования различных событий в системе [15].

Глубокие нейронные сети включают множество слоев и позволяют изучать сложные зависимости в данных системы умного дома. Они могут использоваться для решения различных задач, таких как классификация действий жильцов дома (например, чтение, сон, приготовление пищи) или прогнозирование энергопотребления.

Автокодировщики работают в двух направлениях, выполняя сжатие данных в скрытое пространство и декодирование обратно в исходное пространство. В системах умного дома они могут быть использованы для изучения скрытых структур в данных системы умного дома или для заполнения пропущенных значений в наборе данных [16].

Генеративно-состязательные сети состоят из генератора и дискриминатора, которые соревнуются друг с другом в процессе обучения. Они могут быть использованы для синтеза новых данных системы умного дома, основанных на обучающем наборе данных, или для генерации аномальных данных в целях выявления возможных проблем или атак в системе умного дома [17].

Ассоциативные правила включают:

- алгоритм Apriori;
- алгоритм FP-Growth;
- алгоритм Eclat.

Алгоритм Apriori является одним из наиболее популярных алгоритмов для анализа ассоциативных правил. Он основан на принципе подсчета частоты предметов или их комбинаций и использования этой информации для выявления частых комбинаций. В системах умного дома, данный алгоритм может быть использован для обнаружения частых сочетаний разных устройств или действий жильцов.

Алгоритм FP-Growth используется для извлечения частых комбинаций в наборе данных. Он основан на структуре дерева префиксов, и позволяет эффективно обрабатывать большие наборы данных. В системах умного дома FP-Growth может быть применен для выявления частых сочетаний устройств, событий или состояний.

Алгоритм Eclat (Equivalence Class Clustering and Bottom-Up Lattice Traversal) также используется для анализа ассоциативных правил и выявления частых комбинаций. Он основан на создании классов эквивалентности элементов и их последующей группировки. В системах умного дома Eclat может быть применен для извлечения связей между различными устройствами или состояниями. [18]

III. ОПИСАНИЕ ПРЕДЛАГАЕМОГО ПОДХОДА К АНАЛИЗУ ДАННЫХ СИСТЕМ УМНОГО ДОМА

Предлагаемый подход к анализу данных систем умного дома основан на использовании мобильных

сверточных нейронных сетей MobileNets, которые представляет собой набор моделей компьютерного зрения, оптимизированных для TensorFlow. Эти модели предназначены для достижения высокой точности с использованием ограниченной вычислительной мощности и ограниченных ресурсов, создавая, таким образом, легкие сверточные нейронные сети.

Мобильная CNN-сеть настраивается с помощью следующих двух гиперпараметров: (1) разрешения входного изображения и (2) относительного размера по сравнению с крупнейшей мобильной сетью. Эти гиперпараметры масштабируют соотношение между точностью и задержкой. Логично, что выбор большего разрешения изображения приводит к более трудоемкой, но более точной модели. В соответствии с этим сценарием в предлагаемом подходе сохранены параметры по умолчанию с разрешением входного изображения 224 пикселей и долей 0,5 от модели. Эти 2 параметра были переданы внутри переменных оболочки Linux.

В начале обучения был использован скрипт на Python, полученный из репозитория TensorFlow. Процесс переобучения занял большое количество времени, однако в конце, после проведения анализа всех изображений, вычисления значений узких мест и подачи входных данных на последний этап классификации, вывод сценария показал окончательную точность теста на уровне 91,9 %, что демонстрируется на рис. 1.

```
INFO:tensorflow:Step 3990: Train accuracy = 100.0%
INFO:tensorflow:Step 3990: Cross entropy = 0.003648
INFO:tensorflow:Step 3990: Validation accuracy = 92.0% (N=100)
INFO:tensorflow:Step 3999: Train accuracy = 100.0%
INFO:tensorflow:Step 3999: Cross entropy = 0.003232
INFO:tensorflow:Step 3999: Validation accuracy = 89.0% (N=100)
INFO:tensorflow:Final test accuracy = 91.9% (N=136)
```

Рис. 1. Проверка работы алгоритма

Всего проводится 10 этапов обучения. Для каждого из этапов случайным образом выбирается набор из 10 изображений, которые загружаются в конечный слой для получения прогнозов. Прогнозы впоследствии сравниваются с начальными обучающими метками и, следовательно, обновляются методом обратного распространения. Идея алгоритма обратного распространения состоит в том, чтобы, основываясь на вычислении ошибки, возникшей в выходном слое нейронной сети, пересчитать значение весов последнего слоя нейронов и, таким образом, перейти к предыдущим слоям от начала до конца, то есть обновить все веса слоев от последнего до достижения входного слоя сети, для этого выполняя обратное распространение ошибки, полученной сетью.

Перед выполнением обучения в фоновом режиме запускается средство мониторинга TensorBoard, входящее в состав TensorFlow. Этот процесс выполнялся параллельно.

После выполнения всего процесса обучения был запущен окончательный сценарий для проверки точности оценки модели, который вернул значение 91,9 % (таблица I). Это число отражает общую производительность модели в реальном сценарии классификации, и, поскольку обучение проводилось только по 3 классам, можно было получить высокую точность.

ТАБЛИЦА I. Точность, полученная при обучении

Номер эпохи	Полученная точность	Количество обучающих данных
0	0.435	100
1	0.551	200
2	0.688	300
3	0.745	400
4	0.814	500
5	0.857	600
6	0.873	700
7	0.894	800
8	0.901	900
9	0.909	1000
10	0.911	1100
11	0.914	1200
12	0.918	1300
13	0.919	1400
14	0.919	1500

IV. CONCLUSION

В статье рассматривается подход к эффективной обработке информации с датчиков «умного дома», основанный на совместном использовании архитектуры сверточной нейронной сети MobileNet и программной библиотеки машинного обучения TensorFlow, а также средства мониторинга TensorBoard. Экспериментальные результаты показали, что предложенный подход позволяет достичь показателя точности 91,9% и, соответственно, оперативности обработки информации.

Направление дальнейших исследований связано с повышением точности алгоритма и сокращением времени обучения сверточной нейронной сети.

СПИСОК ЛИТЕРАТУРЫ

- [1] Zantalis F., Koulouras G., Karabetsos S., Kandris D. A review of machine learning and IoT in smart transportation. *Future Internet*. 2019, vol. 11, no. 4, p. 94.
- [2] Malik R.A., Kawoosa A.I., Zargar O.S. Machine learning in the internet of things - standardizing IoT for better learning. *International journal of advance research in science and engineering*. 2018, vol. 7, no. 4, pp. 1676-1683.
- [3] Mahdavinjad M.S., Rezvan M., Barekatin M., Adibi P., Barnaghi P., Sheth A.P. Machine Learning for Internet of Things Data Analysis: A Survey. *Digital Communications and Networks*. 2018, vol. 4, no. 3, pp. 161-175.
- [4] Mao J., Lin Q., Bian J. Application of learning algorithms in smart home IoT system security. *Mathematical Foundations of Computing*. 2018, vol. 1, no. 1, pp. 63-76.
- [5] Biggio B., Fumera G., Roli F. Evade Hard Multiple Classifier Systems. In: *Applications of Supervised and Unsupervised Ensemble Methods*. *Studies in Computational Intelligence*. 2009, vol. 245, pp. 15-38.
- [6] Zhou Z., Firestone C. Humans can decipher adversarial images. *Nature communications*. 2019, vol. 10, p. 1334.
- [7] Zhang Z., Chen M., Backes M., Shen Y., Zhang Y. Inference Attacks Against Graph Neural Networks. 2021, arXiv:2110.02631, 18 p.
- [8] Irfan M.M., Ali S., Yaqoob I., Zafar N. Towards Deep Learning: A Review on Adversarial Attacks. In: *Proceedings of the International Conference on Artificial Intelligence*, 2021, pp. 91-96.
- [9] Papernot N., McDaniel P., Sinha A., Wellman M.P. SoK: Security and Privacy in Machine Learning. In: *Proceedings of the European Symposium on Security and Privacy*, 2018, pp. 399-414.
- [10] Dekel O., Shamir O., Xiao L. Learning to classify with missing and corrupted features. *Machine Learning*. 2010, vol. 81, pp.149-178.
- [11] Goodfellow I., McDaniel P., Papernot N. Making Machine Learning Robust Against Adversarial Inputs. *Communications of the ACM*. 2018, vol. 61, pp. 56-66.
- [12] Liu W., Chawla S. Mining adversarial patterns via regularized loss minimization. *Machine Learning*. 2010, vol. 81, pp. 69-83.

- [13] Li J., Yang Y., Sun J., Tomsovic K.L., Qi H. Towards Adversarial-Resilient Deep Neural Networks for False Data Injection Attack Detection in Power Grids. 2021, ArXiv:2102.09057, 10 p.
- [14] Guo C., Rana M., Cisse M., Maaten L. Countering adversarial images using input transformations. 2017, arXiv:1711.00117, 12 p.
- [15] Li H., Xu X., Zhang X., Yang S., Li B. QEBA: Query-efficient boundary-based blackbox attack. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 1218-1227.
- [16] Li Q., Guo Y., Chen H. Practical no-box adversarial attacks against DNNs. 2020, arXiv:2012.02525, 8 p.
- [17] Li S., Zhu S., Paul S., Chowdhury R., Song C., Krishnamurthy S., Swami A., Chan K. Connecting the dots: Detecting adversarial perturbations using context inconsistency. In: European Conference on Computer Vision, 2020, pp 396-413.
- [18] Dong Y., Pang T., Su H., Zhu J. Evading defenses to transferable adversarial examples by translation-invariant attacks. 2019, arXiv:1904.02884, 13 p.