

УДК 621.324.21

АЛГОРИТМ ГЕНЕРАЦИИ ДАТАСЕТА ДЛЯ ПРИЛОЖЕНИЙ МАШИННОГО ОБУЧЕНИЯ СИСТЕМ V2I

Е. А. Лопухова¹, А. А. Юдин², Р. Р. Алтынбаев³, Е. П. Грахова⁴

¹lopuhova.ekaterina@net.ugatu.su, ²zoomjoker59@gmail.com, ³altynbaev941@gmail.com,
⁴grakhova.ep@ugatu.su

¹⁻⁴ ФГБОУ ВО «Уфимский университет науки и технологий» (УУНИТ)

Поступила в редакцию 25.04.2023

Аннотация. В статье представлен алгоритм генерации датасета, отражающего процесс подключения транспортных средств к сети V2I, для обучения интеллектуальных приложений. Разработана модель, представляющая выделенный участок городской трассы в рамках теории графов и имитирующая движение автомобилей, осуществляющих ретрансляцию своих текущих характеристик. Описана апробация полученного датасета для обучения нейронных сетей и алгоритмов градиентного бустинга в среде Python.

Ключевые слова: V2I, машинное обучение, датасет, нейронные сети, градиентный бустинг.

ВВЕДЕНИЕ

Человеческий фактор является основной причиной ДТП. Если снизить или полностью исключить вероятность проявления неадекватного поведения за рулем, и оставить опыт водителей, обеспечивающий безопасность всех участников дорожного движения, возможно существенно повысить безопасность на дорогах, оптимизировать движение транспортных средств и повысить эффективность использования топлива. Для имитации когнитивных способностей человека в системах помощи водителю и беспилотных автомобилях необходимо применение интеллектуальных алгоритмов, реализующих различные функции мониторинга окружающей среды, состояния транспортного средства, а также самого водителя и управления подсистемами на основе собранной информации.

Для создания комплексного решения для контроля всего дорожного движения, а не отдельного автомобиля, необходима технология беспроводного обмена данными между транспортными средствами и элементами интеллектуальной транспортной системы. С этой целью была создана единая технология связи транспортных средств со множеством объектов (vehicle-to-x, V2X) [1], включающая в себя систему связи между транспортными средствами (vehicle-to-vehicle, V2V), между транспортными средствами и дорожной инфраструктурой (vehicle-to-infrastructure, V2I) и между транспортными средствами и пешеходами (vehicle-to-device, V2D). В технологии V2X интеллектуальные алгоритмы, в частности алгоритмы машинного обучения (ML), являются основными компонентами, регулирующими на основе полученных данных как поведение транспортного средства в целом, так и работу отдельных приложений. В частности, работа интеллектуальных алгоритмов в системах V2X может быть направлена на минимизацию задержек, что не только влияет на удовлетворенность пользователя услугами сети, но и на

безопасность участников дорожного движения путем повышения скорости реакции на динамические изменения внешней среды. Подобные алгоритмы классификации и обучения с подкреплением [2–4] решают задачи определения стратегий распределения ресурсов для каналов V2V и V2I в зависимости от текущей ситуации. В иных случаях минимизация задержек может происходить на этапе установления соединения, где алгоритмы машинного обучения, согласно данным, получаемым из внешней среды, могут с высокой точностью позиционировать объект дорожного движения для регулирования направленности основного излучения антенной системы [5–6].

В свою очередь на этапе моделирования в процессе проектирования схемы работы всех вышеприведенных интеллектуальных приложений V2X необходим датасет, предоставляющий специфичные для конкретного случая применения параметры окружающей среды. Далеко не всегда возможно или экономически оправдано накопление информации от фактической реализации исследуемого процесса, особенно, учитывая необходимость сведения разнородных, но связанных данных. Эту проблему можно обойти путем моделирования изучаемого процесса. Степень приближенности модели к реальности напрямую влияет на приспособленность обученного алгоритма машинного обучения к реальным условиям эксплуатации. Примером разработки специфичного датасета для моделирования распространения сигналов от объектов в системе 5G MIMO является потенциально применимый для приложений глубокого обучения алгоритм генерации данных [7], включающий в себя расчет эволюции каналов во времени.

Целью данной работы является разработка алгоритма генерации датасета для интеллектуальной системы позиционирования транспортных средств, согласно данным, получаемым от них, и априорным данным о контролируемой зоне обслуживания базовой станции.

Современные алгоритмы для определения параметров наведения и удержания луча опираются прежде всего на данные спутниковых систем навигации. Однако в ряде случаев они могут быть недоступны или же задержки при их передаче к объектам системы V2X могут быть велики, например, в условиях неустойчивого соединения. Следовательно, для алгоритмов машинного обучения в данных условиях необходимы иные методы позиционирования объекта отслеживания. Априорные знания о зоне обслуживания базовой станции, где располагается интеллектуальный алгоритм, могут существенно снизить степень неопределенности о местоположении автомобиля, обладающего определенными характеристиками.

РАЗРАБОТКА АЛГОРИТМА ГЕНЕРАЦИИ ДАТАСЕТА

Входные данные предоставлялись алгоритму машинного обучения посредством ретрансляции объектами сети V2I своих постоянных или измеряемых датчиками параметров. К наиболее важным для установления устойчивого соединения данным были отнесены временные метки (моменты отправки сообщений), вид транспортного средства, позволяющий определить упреждение, опираясь на значение его средней скорости на трассе, и уровень принятого сигнала, определяемый расстоянием до объекта и профилем трассы между ним и базовой станцией.

Информацию о предполагаемом нахождении автомобиля и его поведении на трассе предоставляли цифровые модели рельефа и трафика, известные обучаемому алгоритму. Одним из способов представления цифровой модели трафика является применение теории направленных графов, как показано на рис. 1. Каждая дорожная полоса была представлена направленным ребром графа, а повороты – его вершинами. С учетом того, что запрос на подключения к сети осуществлялся только в процессе передвижения по трассе, были заранее заданы возможные маршруты передвижения автомобилей, а угол поворота относительно прямолинейного движения был введен как дополнительный информационный параметр. Подобную информацию можно получить от датчика поворота руля. Таким образом, особыми требованиями для функционирования автомобиля в моделируемой системе является наличие датчика поворота руля и системы синхронизации времени.

В сообщение-запрос транспортного средства для установления устойчивого соединения входили разница во времени между моментами ретрансляции запроса о подключении к сети и принятия сообщения (определяемую по меткам времени), тип транспортного средства (легковой и грузовой автомобиль) и уровень сигнала на приеме. Информация из сообщения-запроса являлась внешней относительно интеллектуальной системы, остальные входные показатели фиксировались после получения сообщения с учетом априорных данных. В простейшем случае при исключении возможности многолучевого распространения расчет временного сдвига производился исходя из расстояния в момент трансляции сообщения между объектом и базовой станцией. Упреждение рассчитывалось на основе класса транспортного средства, характеризующего его среднюю скорость на обслуживаемом участке дороги.

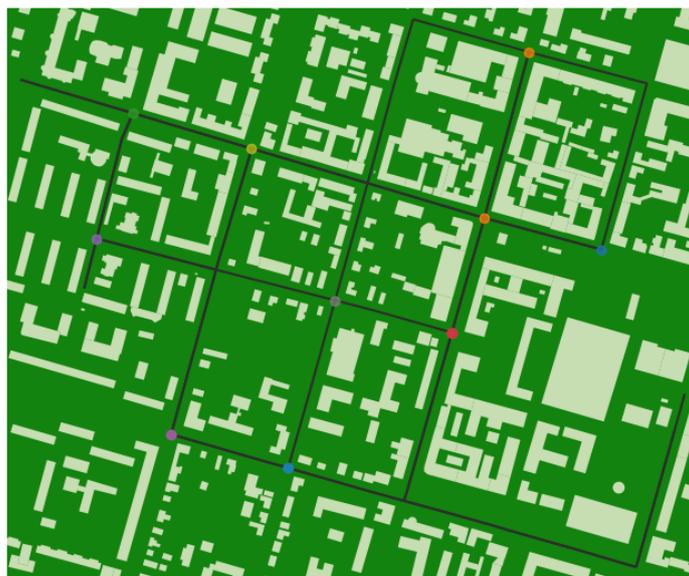


Рис. 1. Направленный граф выделенного дорожного участка

Для учета потерь в свободном пространстве на участке обслуживания была применена модель Лонгли-Райса, предоставляющая достаточную точность относительно реальных измерений напряженности поля [8]. Модель распространения учитывает особенности профиля трассы между приемником и передатчиком, что предоставляет значительно больше информации нежели затухание сигнала в свободном пространстве. При расчете модели распространения используются теоретические методы отражения от неровной поверхности, преломления через атмосферу, дифракции вокруг Земли и на острых препятствиях, а также рассеяния в тропосфере. Они объединяются с помощью эмпирических соотношений, полученных в результате подгонки к измеренным данным.

Информация от модели прогнозирования зон радиообслуживания позволила сгруппировать данные о схожих уровнях сигнала на входе антенной системы, которые в сочетании с другими параметрами и априорными знаниями о зоне обслуживания дают возможность локализовать местонахождение искомого объекта. При генерации датасета имитировалась посылка запросов в произвольных точках трассы в пределах зоны обслуживания базовой станции. На основании этих данных вычислялся номер сектора антенной системы, в пределах которого должно быть транспортное средство на момент принятия запроса относительно предполагаемого размещения базовой станции. Антенная система была представлена фазированной антенной решеткой с углом сканирования 120° и шириной бокового лепестка 9° , следовательно, количество секторов направления основного излучения 11. Задачей системы являлось определение сектора излучения к транспортному средству по входным данным. Примеры реализаций сформированного датасета представлены в табл. 1.

Примеры входных данных и их меток из сформированного датасета

№	Временные задержки, мкс	Угол поворота	Тип транспортного средства	Уровень принятого сигнала, дБ	Номер сектора
1	1,3549	0,0	Грузовой	-113,4960	4
2	4,0122	90,6	Легковой	-118,4307	7

АПРОБАЦИЯ СФОРМИРОВАННОГО ДАТАСЕТА

Апробация сформированного датасета, показывающая, что вектора входных значений обладают достаточным набором скрытых признаков для определения выходного пространства состояний, была проведена на нейронных сетях (НС) и алгоритме градиентного бустинга. Градиентный бустинг представляет собой ансамблевую регрессионную модель [9]. Обучение модели осуществляется путем последовательного приближения ошибки предсказания от ансамбля базовых деревьев на предыдущей итерации к минимуму методом наименьших квадратов на каждой итерации. Окончательная модель представляет собой сумму исходного дерева и последующих деревьев, обученных по цепочке значений функции ошибки. НС являются достаточно хорошим решением для однородных данных, но на разнородных демонстрируют более низкие результаты в сравнении с ансамблевыми методами, что наглядно продемонстрировано на рис 2. Помимо алгоритмов градиентного бустинга XGBoost, была протестирована также работа сетей с 2 скрытыми слоями FC_Conv1d_FC – сверточной НС, FC_LSTM_FC – рекуррентные НС и FC_Linear – полносвязная НС.

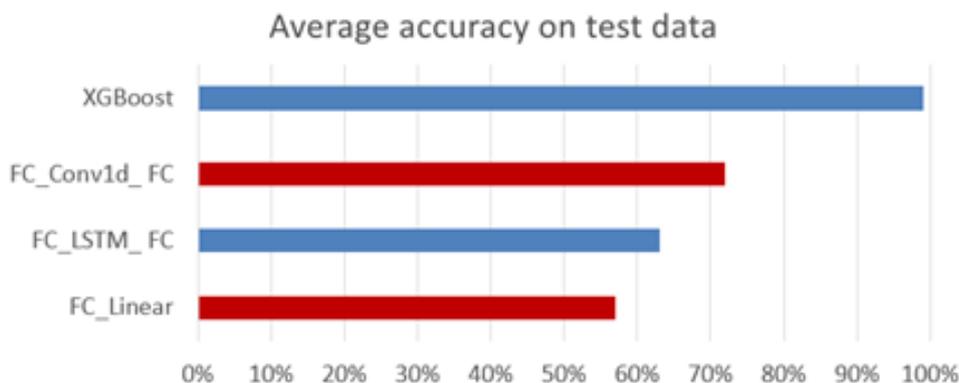


Рис. 2. Точность алгоритмов ML на сформированном датасете

Для анализа логики работы алгоритма XGBoost была построена гистограмма важности входных параметров, зависящая от частоты присутствия признака в качестве узла, решающего правила в итоговом дереве принятия решений, представленная на рис. 3. Из гистограммы видно, что основными информационными параметрами являются временная задержка и уровень сигнала. Они позволяют кластеризовать всю обслуживаемую область исходя из дальности распространения сигнала и особенностей профиля трассы между определяемым сектором и базовой станцией. Остальные параметры играют роль уточняющих положение в секторе транспортного средства на момент установления устойчивого соединения.

Результаты моделирования показали требуемые задержки, приведенные в табл. 2, на обучение и обработку входного блока информации представленных алгоритмов машинного обучения. Полученные данные свидетельствуют о преимуществе градиентного бустинга среди прочих алгоритмов в скорости обучения и обработки информации.

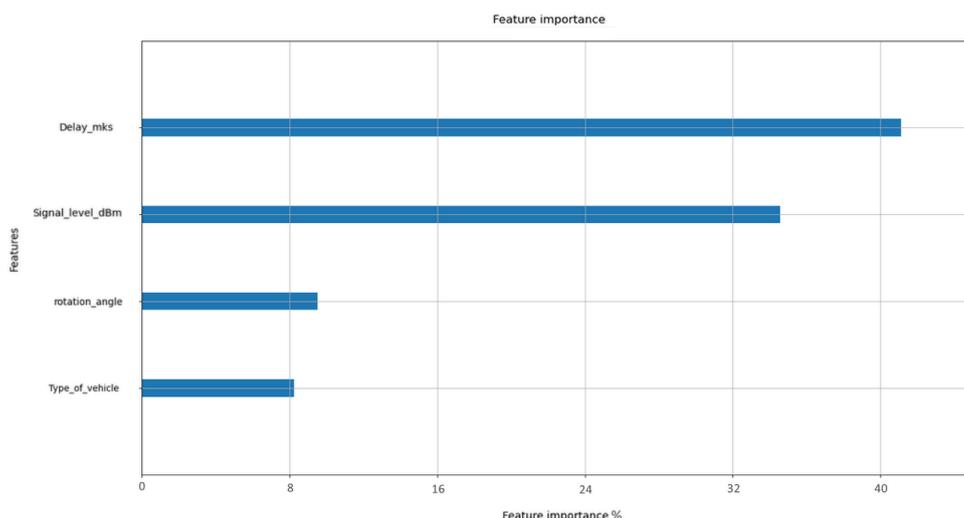


Рис. 3. Важность признаков датасета в итоговом дереве принятия решений

Таблица 2

Время обучения и отклика алгоритмов машинного обучения на сформированном датасете

Алгоритм	XGBoost	FC_Convld_FC	FC_LSTM_FC	FC_Linear
Время обучения, с	11,01	23,60	29,70	16,10
Время отклика, мс	0,056	13,100	11,400	4,500

ЗАКЛЮЧЕНИЕ

В статье приведено приложение для генерации датасета для обучения алгоритмов машинного обучения в системах V2I и его апробация на нейросетевых структурах и алгоритме градиентного бустинга. Полученные данные свидетельствуют о присутствии в структуре сформированных данных параметров, позволяющих интеллектуальным алгоритмам с достаточной точностью отобразить входные данные в выходное пространство. Апробация показала, что сформированный датасет наиболее приспособлен для применения с ансамблевыми алгоритмами, работающими с разнородными данными. Для дальнейшего развития данной темы необходимо моделирование работы V2I, более приближенное к реальным процессам, где учитываются искажение сигнала в условиях многолучевого распространения и более сложное поведение транспортных средств.

СПИСОК ЛИТЕРАТУРЫ

1. **Парамонов М. И.** О процессах в нелинейных реактивностях, провоцирующих возникновение и накопление энергии при параметрическом резонансе// European Science. 2014. № 1.
2. **Агафонов А. И.** Параметрический трансформатор/Патент RU №2040058, МПК H01F27/28, G05F3/06, опубли. 20. 07. 1995.
3. **Стыскин А. В., Уразбахтина Н. Г., Терликова Н. С.** Параметрический трансформатор/Патент RU 116268 U1, МПК H01F 19/04, H01F 27, опубли. 20. 05. 2012.
4. **Парамонов М. И.** Патент на изобретение №2674009 RU, МПК H01F 29/14, опубли. 12. 04. 2018// Патент RU 116268 U1.

ОБ АВТОРАХ

ЛОПУХОВА Екатерина Александровна, аспирант 1-го курса, инженер каф. ТС факультета АВИАЭТ.

ЮДИН Артем Александрович, студент 3-го курса каф. АСУ ФИРТ.

АЛТЫНБАЕВ Радмир Рустемович, студент 3-го курса каф. АСУ ФИРТ.

ГРАХОВА Елизавета Павловна, к.т.н., заведующая НИЛ «Сенсорные системы на основе устройств интегральной фотоники», старший научный сотрудник, доцент каф. ТС факультета АВИАЭТ.

METADATA

Title: Dataset generation algorithm for machine learning applications in V2I systems

Authors: ¹E. A. Lopuhova, ²A. A. Yudin, ³R. R. Altinbaev ⁴E. P. Grakhova

Affiliation:

¹⁻⁴ Ufa University of Science and Technology (UUST), Russia.

E-mail: ¹lopuhova.ekaterina@net.ugatu.su, ²zoomjoker59@gmail.com, ³altyn-baev941@gmail.com, ⁴grakhova.ep@ugatu.su

Language: Russian.

Source: Molodezhnyj Vestnik UGATU (scientific journal of Ufa University of Science and Technology), no. 3 (29), pp. 115–120, 2023. ISSN 2225-9309 (Print).

Abstract: The paper presents an algorithm for generating a dataset, reflecting the process of connecting vehicles to the V2I network, for training intelligent applications. A model was developed, which represents a selected segment of a city highway in the framework of graph theory and simulates the movement of cars, transmitting their current characteristics. We describe the testing of the obtained dataset for training neural networks and gradient boosting algorithms in Python code.

Keywords: V2I; machine learning; dataset; neural networks; gradient boosting.

About the authors:

LOPUHOVA Ekaterina Alexandrovna, 1st year postgraduate student, engineer at Ufa state aviation technical university.

YUDIN Artem Aleksandrovich, 3rd year student, Ufa University of Science and Technology (UUST).

ALTINBAEV Radmir Rustemovich, 3rd year student, Ufa University of Science and Technology (UUST).

GRAKHOVA Elizaveta Pavlovna, Head of research laboratory, Senior Researcher, Associate Professor, Ufa University of Science and Technology (UUST).