

Методы глубокого обучения в прогнозировании трафика в сетях автотранспорта на основе распределенной вычислительной системы

Т. В. Лаптева

Санкт-Петербургский государственный университет
телекоммуникаций им. проф. М.А. Бонч-Бруевича

1978.laptevat.v@mail.ru

А.С.А. Мутханна

Санкт-Петербургский государственный университет
телекоммуникаций им. проф. М.А. Бонч-Бруевича

ammarexpress@gmail.com

Аннотация. Динамичность сетевой инфраструктуры, рост количества сетевых устройств ведут к увеличению потоков служебных данных. Причем для каждого вида услуг предъявляются разные требования к качеству обслуживания (QoS). Ранее применяемые инструменты инженерии управления телекоммуникационным трафиком становятся ограниченными. Использование алгоритмов машинного обучения и технологий искусственного интеллекта с интеграцией беспилотных летательных аппаратов (БПЛА) в общую структуру современных сетей для сетевой поддержки предоставляет возможность эффективного управления потоками данных.

Ключевые слова: граничные вычисления, машинное обучение, искусственный интеллект, беспилотные летательные аппараты

I. ВВЕДЕНИЕ

Стремительное увеличение объемов трафика данных в сетях связи общего пользования повышает требования к скорости их обработки и качеству передачи. Появление новых типов оборудования, значительный рост количества устройств IoT, перспектива развития сетей автономных транспортных средств и прочие нововведения требуют минимизации времени круговой задержки до 1 м/с, повышения надежности сети и качества обслуживания. В связи с этим должны делаться ставки на технологии управления сетью связи, позволяющие проводить «умный» мониторинг в условиях динамичности сетевой инфраструктуры, в целях своевременного выявления узких мест, и оперативную ее переконфигурацию с учетом требований сервисов.

Решение таких задач под силу интеллектуальной распределенной архитектуре. Концепция программно-конфигурируемых сетей (SDN) в связке с виртуализацией сетевых функций (NFV), обозначенные пунктом 2.3.2 рекомендаций Международного союза электросвязи М.2083-0, обеспечивают управление сетью, а в сочетании с технологиями граничных распределенных вычислений (MEC) позволяют повысить ее надежность и масштабируемость. Обмен данными между устройствами по технологии прямой связи (D2D), минуя базовую станцию или базовую сеть, снижает нагрузку на сеть и сокращает задержку. Сетевая архитектура с предлагаемой структурой позволяет приблизиться к требованиям высокой надежности сети и ультрамалых задержек.

"Научная статья подготовлена в рамках прикладных научных исследований СПбГУТ, регистрационный номер 1023031600087-9 в ЕГИСУ НИОКТР."

Динамичность сетевой инфраструктуры, рост количества сетевых устройств ведут к увеличению потоков служебных данных. Причем для каждого вида услуг предъявляются разные требования к QoS. Ранее применяемые инструменты инженерии управления телекоммуникационным трафиком становятся ограниченными. Использование алгоритмов машинного обучения (ML) и технологий искусственного интеллекта (AI) предоставляет возможность эффективного управления потоками данных.

II. ПЕРСПЕКТИВЫ ИСПОЛЬЗОВАНИЯ МАШИННОГО ОБУЧЕНИЯ И ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В СЕТЯХ СВЯЗИ

В глобальном плане в условиях увеличивающегося потока трафика коммуникаций между машинами в сетях связи должны быть реализованы требования к сверхнадежным сетям обеспечивающих задержку до 1 м/с. (URLLC) [1]. В том числе должна быть разработаны способы контроля и возможность управления всеми современными и разрабатываемыми в перспективе услугами, к которым предъявляются жесткие требования по качеству обслуживания и восприятия для конечного пользователя услуг.

Услуги и сервисы, существующие и новые, активно предлагаемые, и внедряемые в повседневную жизнь, задают высокие требования к производительности и отказоустойчивости сетей связи. Применение новейших технологий на всех уровнях сетей, расширение масштабируемости, технологии облачных, а также туманных вычислений позволяют не в полной мере реализовать предъявляемые требования. Важно выполнить задачи в части улучшения и оптимизации логики мониторинга, обработки и управления сетевым трафиком. Это решение возможно за счет внедрения новейших методик AI.

Алгоритмы ML и глубокого обучения (DL) возможно применять для решения вопросов разгрузки: сокращения вычислений и энергопотребления, для минимизации задержки и времени отклика, а также управления распределением ресурсов. ML позволяет получить более точные и актуальные результаты при обработке больших объемов данных, а использование сетей 5G позволяет гарантированно осуществлять передачу спрогнозированной информации, что позволяет алгоритмам машинного обучения оптимально функционировать. DL – это один из перспективных методов машинного обучения, отличающийся от традиционных методов свойством эффективно

обрабатывать нелинейные данные, что также обеспечивает надежные возможности анализа, обучения, оптимизации и интеллектуального распознавания входных данных.

III. ПРОГНОЗИРОВАНИЕ И УПРАВЛЕНИЕ ПОТОКАМИ ТРАФИКА С ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

В современных условиях применение AI осуществляется преимущественно по двум направлениям: прогнозирование и распознавание. К примеру: распознавание атаки на контроллер сети, или распознавание типа трафика. В целом, с учетом множества критериев, осуществляется прогнозирование нагрузки контролируемого потока данных на инфраструктуру оператора связи. [2], [3].

В данной концепции применения AI в сетях связи особенно важен принцип микросервисного подхода. При таком подходе определенные функции машинного обучения выделяются в независимые модули, которые взаимодействуют между собой и вышестоящими управляющими модулями. При использовании микросервисного подхода появляется гибкость в управлении системой, и при необходимости, ее обновлении. Кроме этого, появляется независимость в масштабировании и простота в эксплуатации. Например, для увеличения производительности или при возникновении необходимости обеспечить новые системы интеллектуальным модулем управления, достаточно размножить соответствующие микросервисы и произвести их миграцию в соответствующий сегмент сети [4].

IV. АЛГОРИТМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ПРОГНОЗИРОВАНИЯ ТРАФИКА

Прогнозирование информации, основанное на связи между элементами последовательности, позволяет выявлять нелинейные зависимости и работать со сложными динамически меняющимися данными большого объема, обеспечивает безопасность, надежность и связность выбранных сетей. DL является одним из перспективных методов машинного обучения, позволяющий эффективно обрабатывать нелинейные данные. За счет итерационного процесса быстрого построения сложных статистических моделей этот метод позволяет создать точные прогнозные модели на основе больших объемов немаркированных и не структурированных данных.

Модели на основе нейронных сетей (NN), такие как: нейронная сеть с прямой связью, нечеткая нейронная сеть, рекуррентная нейронная сеть (RNN), Гауссовский процесс и гибридная нейронная сеть давно и успешно используются для задач прогнозирования трафика.

Прогнозирование временных рядов строится на основе предыдущей и текущей информации и относится к прогнозированию следующих значений системы. Обычно для получения ожидаемых данных используется предопределенное вычислительное моделирование (или гибридное моделирование). Для этого используют рекуррентные нейронные сети с памятью, которые могут предсказывать цели, основываясь на последовательных данных в зависимости от времени.

V. ОБЗОР РАБОТ МЕТОДОВ ГЛУБОКОГО МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОПТИМИЗАЦИИ РЕСУРСОВ СЕТИ

В настоящее время алгоритмы глубокого обучения широко используются во многих сферах жизни. Например, таких как обработка естественного языка, игры, Интернет Вещей, компьютерное зрение, распознавание речи и т. д. [4], [5].

Для решения различных задач при исследованиях автомобильных периферийных вычислительных систем адаптация моделей AI осуществляется опытным путем с использованием обучения с подкреплением [6].

Этот метод был предложен для децентрализованного распределения ресурсов при обработке ограничений задержки коммуникаций между транспортными средствами [7].

В работе [8] совместно рассматривались коммуникационные и вычислительные ресурсы для планирования передачи данных, где пропускная способность данных и вычислительная мощность при движении могли бы быть предоставлены транспортным средствам через придорожный блок около дороги. В последствии вычисления, кэширование, коммуникации и совместные вычисления были сформулированы как задачи оптимизации, чтобы минимизировать затраты на обработку данных в условиях ограничений по задержке. Наконец, для выработки оптимальной стратегии планирования данных было использовано глубокое обучение с подкреплением. Однако большинство из этих работ касается одноуровневой среды, которая может вызвать перегрузку и привести к более длительной задержке обработки при вычислении нагрузки, если большое количество транспортных средств загружает свои вычислительные задачи на один и тот же сервер. Кроме того, использование традиционных методов обучения с подкреплением считается непомерно сложными с точки зрения вычислений для крупномасштабных сред.

VI. ПРИМЕНЕНИЕ СОСТАВНОЙ МОДЕЛИ GRU-BiLSTM НА ОСНОВЕ DL В ПРОГНОЗИРОВАНИИ ТРАФИКА С ИНТЕГРАЦИЕЙ БПЛА В ОБЩУЮ СТРУКТУРУ СОВРЕМЕННЫХ СЕТЕЙ ДЛЯ СЕТЕВОЙ ПОДДЕРЖКИ

Концепция возможностей IoT подразумевает сценарии высокоплотного и сверхплотного построения сетей связи. В свою очередь, развитие сетей с ультрамалыми задержками – как одного из ключевых направлений развития сетей пятого поколения, направлены на реализацию концепции автономных транспортных средств (AVs). Использование технологий МЕС с поддержкой беспилотных летательных аппаратов (UAVs) и методов машинного обучения для прогнозирования сетевого трафика в области сетей и систем связи позволит найти пути решения задач по построению таких сетей.

Достижение цели возможно за счет разработки модели сети, в которой на уровне облака МЕС используются методы глубокого обучения для прогнозирования и извлечения скорости движения на основе AV-системы. Система состоит из двух частей:

- модуль извлечения информации о трафике на основе AVs;

- модуль прогнозирования скорости трафика на основе глубокого обучения.

AV-система моделируется в среде Python для создания набора обучающих данных о дорожном движении. Для прогнозирования трафика на основе модели DL используется среда Matlab. Собранный и обработанный набор обучающих данных загружается в модель DL и разделяется на 70 % для обучения и 30 % для тестирования.

В прогнозировании потоков трафика для повышения точности при одновременном сокращении времени предлагается составная модель DL, объединяющая закрытый регулируемый рекуррентный модуль (GRU) и расширение долгой краткосрочной памяти с двунаправленным потоком информации (BiLSTM). Она сочетает в себе преимущества GRU и BiLSTM с точки зрения эффективности и точности прогнозирования для использования в AV-трафике в густонаселенных районах.

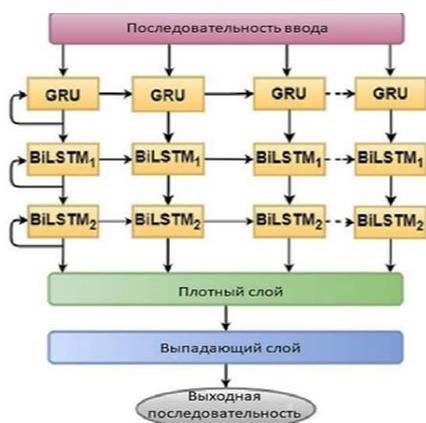


Рис. 1. Предлагаемая гибридная модель GRU-BiLSTM

На рис. 1 изображена сетевая структура модели интегрированного прогнозирования. Модель прогнозирования передает в модель DL параметры пропускной способности AV-системы. Вначале используется уровень GRU, который обучает данные с базовой сетевой структурой, т. е. меньшим количеством параметров и более простой сходимостью. Обучение модуля GRU происходит быстро и эффективно; однако его прогноз не так точен, как прогноз BiLSTM. Затем модель накладывает два слоя BiLSTM с большим количеством параметров, поскольку двухслойный BiLSTM имеет более высокую точность прогнозирования, чем однослойный. После этого, модель прогнозирует будущий трафик, учитывая время и точность прогнозирования. Многоуровневая структура GRU-BiLSTM состоит из множества уровней, каждый из которых включает один уровень GRU и два уровня BiLSTM. Вместо одного значения уровень BiLSTM отправляет последовательность значений на базовый уровень BiLSTM, по одному для каждого входного временного шага, а не по одному для всех входных временных шагов.

Прогнозирование временных рядов сетевого трафика использует исторические данные с отметками времени для создания научных прогнозов. Модели создаются посредством исторических исследований для проведения наблюдений и принятия стратегических решений. Это может служить упреждающим подходом к задачам

планирования и управления сетью. Прогнозирование транспортного потока важно для автономных транспортных средств, чтобы определять маршруты транспорта и принимать адаптивные решения на основе окружающих объектов. Для удовлетворения будущих потребностей сети требуется прогнозирование их параметров.

Предлагаемая модель летающего тумана с помощью AV-системы, представленная на рис. 2, включает в себя уровень наземного сегмента и уровень летающего тумана и используется для создания набора обучающих данных DL. В данной модели сети серверы мобильных граничных вычислений размещаются на БПЛА, что обеспечивает сокращение времени задержки и энергопотребления. Процедура выгрузки трафика осуществляется в три этапа. Программный профилировщик, размещенный на оконечных устройствах, оценивает сложность выполняемой задачи. По результатам оценки запускается алгоритм принятия решения, который принимает решение о необходимости выгрузки трафика [9].

Слой тумана имеет узлы UAV-Fog с ретрансляторами, маршрутизаторами, точками доступа, базовыми станциями, шлюзами и серверами тумана. Туманные узлы расположены на расстоянии одного шага от края сети. Туманные узлы обычно располагаются между облачными центрами и терминальными устройствами. Уровень тумана в качестве промежуточного уровня повышает качество обслуживания сети за счет хранения и быстрой обработки данных, балансировки нагрузки и низкой задержки.

Уровень тумана состоит из группы облачков установленных на дронах. Эти дроны подключаются специально для достижения целей высокого уровня и служат для передачи вычислительных задач с терминальных устройств на серверы тумана. Устройства воздушного сегмента – это своего рода мобильные серверы граничных вычислений. Они являются неким шлюзом, на котором может проводиться обработка полученных задач, либо при необходимости переносится на более оптимальные для этих целей граничные узлы.

Верхним слоем архитектуры тумана на базе БПЛА является уровень облака. Этот уровень позволяет наземным пользователям получать доступ к ресурсам и сохранять данные в качестве резервной копии. Облачный уровень взаимодействует с наземным сегментом для управления ресурсами и обработки данных. На этом уровне работают центры обработки данных и серверы. Облако предлагает мощные вычислительные возможности и возможности хранения данных, доступные с любого сетевого устройства в любое время. Для обеспечения безопасности этому уровню требуется шифрование с открытым ключом.

В большинстве архитектур Fog на базе AVS контроллер собирает данные о состоянии системы от пользователей, дронов и серверов Fog. Этот контроллер отвечает за управление операциями по разгрузке с помощью методов искусственного интеллекта. Дрон принимает выгруженные задачи из системы IoT и возвращает результаты после выполнения задачи. БПЛА может отправлять данные на близлежащие наземные серверы, если встроенное облако не может справиться со сложной обработкой. Сеть может использовать большое количество БПЛА для покрытия больших площадей при

использовании устройств IoT, таких как телефоны, датчики, автомобили и роботы. AI используются для анализа и обработки информации, предоставляемой пользователями через встроены облака.

Прогнозирование сетевого трафика повышает безопасность дорожного движения, уменьшает заторы и улучшает производительность транспортных сетей.

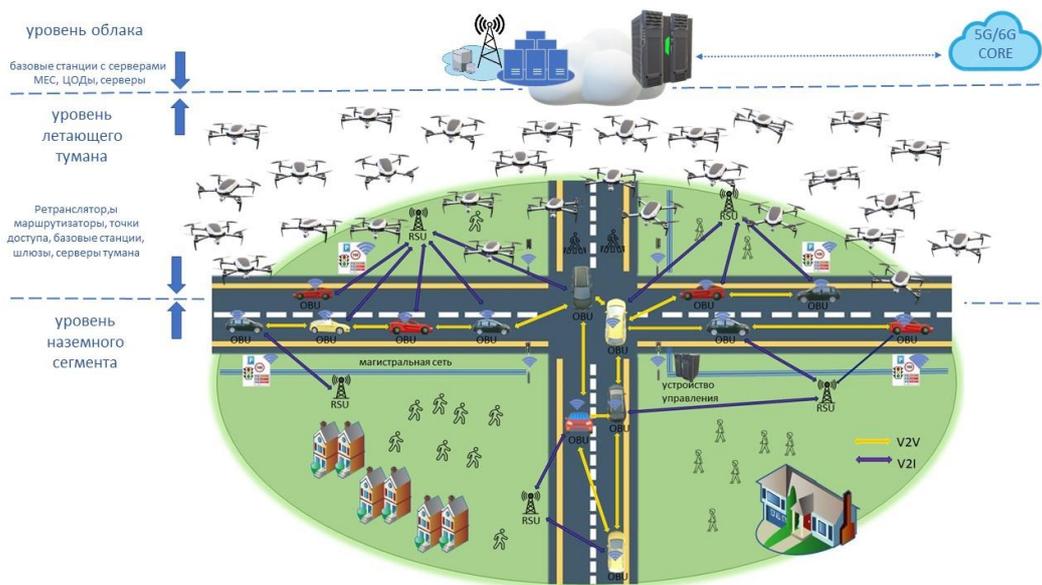


Рис. 2. Архитектура Fog с поддержкой БПЛА

Традиционные методы прогнозирования редко могут уловить и не могут справиться со сложностью и разнообразием моделей трафика. ML может прогнозировать и обнаруживать шаблоны трафика, используя большие и сложные наборы как об оперативных, так и исторически накопленных данных. Данные о сетевом трафике содержат информацию о потоке трафика, его плотности, задержке, потреблении энергии, потере пакетов, пропускной способности, временном интервале и т. д. Системы ML могут анализировать эти данные, чтобы распознавать ключевые проблемы, влияющие на структуру трафика, что идеально подходит для его прогнозирования. Модели машинного обучения прогнозируют будущие модели трафика на основе предыдущих данных о трафике.

VII. ЗАКЛЮЧЕНИЕ

В настоящей статье была рассмотрена гипотеза о том, что глубокая иерархическая модель DL может представлять некоторые функции экспоненциально более эффективно, чем традиционные модели более точно подтверждая ее статус DL. Успех этого подхода в решении сложных задач прогнозирования обычно объясняется глубиной нейронной сети. Поддержка сети с использованием AVS и БПЛА для разгрузки сети и увеличения площади покрытия в сочетании с гибридными составными моделями глубокого обучения позволяет повысить точность и эффективность прогнозирования трафика.

СПИСОК ЛИТЕРАТУРЫ

[1] Кучерявый А. Е. Сети связи с ультрамалыми задержками // Труды НИИР. 2019. № 1. С. 69–74.
 [2] Бородин А.С. Сети связи пятого поколения как основа цифровой экономики / А.С. Бородин, А.Е. Кучерявый // Электросвязь. 2017. № 5. С. 45-49.

[3] Бородин А.С. Искусственный интеллект в сетях связи пятого и последующих поколений / А.Н. Волков, А.С. Мутханна, А.Е. Кучерявый // Электросвязь №1. 2021. С.17-22.
 [4] Хайят М. Усовершенствованная разгрузка вычислений на основе глубокого обучения для многоуровневых автомобильных сетей Edge-Cloud / М. Хайят, И.А. Эльгенди, А.С.А. Мутханна, А.С. Альшахрани, С Альхарби, А. Кучерявый // IEEE Access. 2020. Текст: электронный. – URL: <https://ieeexplore.ieee.org/abstract/document/9146836> (дата обращения 05.11.2023)
 [5] Chen J. Deep learning with edge computing: A review / J. Chen, X. Ran // Proceedings of the IEEE, vol. 107, no. 8, – Aug. 2019. Текст: электронный. – URL: <https://ieeexplore.ieee.org/document/8763885> (дата обращения 05.11.2023)
 [6] Qiao G. Deep reinforcement learning for cooperative content caching in vehicular edge computing and networks / G. Qiao, S. Leng, S. Maharjan, Y. Zhang, N. Ansari // IEEE Internet Things J., vol. 7, № 1, – Jan. 2020. Текст: электронный. – URL: <https://ieeexplore.ieee.org/abstract/document/8879573> (дата обращения 05.11.2023)
 [7] Ye H. Deep reinforcement learning based resource allocation for V2V communications / H. Ye, G.Y. Li, B.-H.-F. Juang // IEEE Transactions on Vehicular Technology, vol. 68, № 4. – Apr. 2019. Текст: электронный. – URL: <https://ieeexplore.ieee.org/document/8633948> (дата обращения 05.11.2023)
 [8] Luo Q. Collaborative data scheduling for vehicular edge computing via deep reinforcement learning / Q. Luo, C. Li, T. H. Luan, W. Shi // IEEE Internet of Things Journal, – Mar. 2020. Текст: электронный. – URL: <https://ieeexplore.ieee.org/document/9047880> (дата обращения 05.11.2023)
 [9] Мутханна А.С.А. Разработка и исследование комплекса моделей и методов интеграции граничных вычислений в сетях связи пятого и шестого поколений специальность 2.2.15 «Системы, сети и устройства телекоммуникаций» диссертация на соискание ученой степени доктора технических наук / Мутханна Аммар Салех Али; Санкт-Петербургский Государственный университет телекоммуникаций. СПб.: 2023 Текст: электронный. – URL: https://www.sut.ru/new_site/images/blocks/1694689561.pdf (дата обращения 05.11.2023)