

УДК 519.688

ИСПОЛЬЗОВАНИЕ МУЛЬТИАГЕНТНОГО ПОДХОДА ПРИ РЕШЕНИИ ЗАДАЧИ ВЫДЕЛЕНИЯ ЭЛЕМЕНТАРНЫХ ОПЕРАЦИЙ НАД ОБЪЕКТАМИ В СИСТЕМЕ ВИДЕОАНАЛИТИКИ

Е.И. Мосейко

Санкт-Петербургский государственный университет
Россия, 199034, Санкт-Петербург, Университетская наб., д. 7–9
E-mail: e.moseyko@spbu.ru

О.Н. Граничин

Институт проблем машиноведения Российской академии наук
Россия, 199034, Санкт-Петербург,
E-mail: oleg_granichin@mail.ru

Ключевые слова: мультиагентная система, видеоаналитика, выделение признаков, классификация действий.

Аннотация: В работе представлен новый алгоритм классификации действий, совершаемых человеком или механизмом над объектами на основе мультиагентного подхода. Мультиагентная система осуществляет верификацию и классификацию действий с последующим разделением сложного графа процесса, совершаемого над объектом, на элементарные компоненты-действия, автоматически пополняя базу знаний. В разрабатываемой системе каждый элемент-действие опознается своим агентом-детектором. Агенты могут кластеризоваться по признакам схожести, найденным самим алгоритмом, тем самым гарантируя сохранение нового неопознанного действия как кандидата на неизвестное элементарное. Онтология агентов (API) на основе json-файлов такова, что технически задача может быть решена иерархически. Т.е. сначала контролируемо обучаемая сеть находит подходящий кластер агентов-детекторов, а затем обрабатывает алгоритм принятия решения о способе классификации и снова происходит поиск очередного уточняющего кластера. Иерархическая система обработки работает на коротком видеофрагменте с трековой информацией или на последовательности фотокадров с дополнительной информацией от сенсоров современного смартфона.

1. Введение

Задача распознавания и классификации действий над объектами стоит уже более 20 лет. Особой популярностью пользуется задача распознавания действий человека, ей посвящено огромное количество исследований во всем мире, т. к. умение максимально точно формализовать действия человека или животного на современном уровне техники позволяет достаточно качественно и дешево создавать роботов, выполняющих рутинную работу.

Как только действие удалось формализовать его тут же перекладывают на «железо». Тысячи работ и образцов кода алгоритмов ежегодно появляются в открытом доступе, количество и качество растет ежемесячно.

Несмотря на такой огромный интерес к данной тематике в промышленности до сих пор нет достаточно универсальной системы видеоаналитики, позволяющей с приличной степенью достоверности определять, выполняет ли исполнитель сложные регламентные действия и оказывать ему информационную поддержку с целью повышения качества работ и снижения издержек, возникающих при ошибках.

Основной причиной, по которой развитие систем видеоаналитики в промышленной безопасности по сути остановилось на определении фактов ношения средств индивидуальной защиты (СИЗ) персоналом и контроля так называемых опасных зон, является очень высокая стоимость кастомизированной разработки систем контроля.

В настоящее время большинство систем использует предварительно обученные модели нейронных сетей (как правило CNN/RNN), но даже при использовании мобильных версий сетей, внедрение алгоритмов, основывающихся на композиции детекторов различных типов сталкивается с огромным количеством технических трудностей. В первую очередь это связано с высокой вычислительной нагрузкой, создаваемой почти любой сетью. Далее, большинство современных программных комплексов не готово к работе с распределенной нагрузкой различного типа, которая никак не формализована. Системы мониторинга направляют данные, получаемые стандартными SDK от производителей камер на общепринятые контейнеры (например, NVIDIA Triton Server), обрабатывающие кадры с захвата стандартным набором алгоритмов, оформленных в виде keras или onnx моделей, затем результат детекции возвращается в виде числового значения в базу данных и оттуда поступает в приложение, обычно называемое «АРМ» (автоматизированное рабочее место оператора). На все эти манипуляции обычно требуется достаточно мощный сервер с GPU, большим объемом оперативной памяти и дискового пространства.

Главная сложность, возникающая при усложнении наблюдаемого процесса, заключается даже не в том, что контейнеры, работающие с данными, приходится запускать параллельно, в лучшем случае применяя к ним алгоритмы балансировки загрузки. Основной проблемой является сложность сбора датасетов, хранения и первичной обработки (разметки) данных предметной области.

В последнее десятилетие стало модным использовать слова мультиагентная система и глубокое обучение. По всему миру объявляются конкурсы по созданию эффективных алгоритмов, моделей, фреймворков для решения сложных задач распознавания. В последних работах одним из важных направлений являются исследования и обзоры по мультиагентным алгоритмам, где агентами являются объекты распознавания или вычислители, на которых работают детекторы.

Наша работа посвящена подходу, в котором сама система подготовки данных становится мультиагентной средой, а агентами являются элементарные алгоритмы-детекторы, которые умеют проявлять групповой интеллект, а именно кластеризоваться, образуя тем самым дерево признаков (некоторую систематизацию).

2. Постановка задачи

В самом широком смысле требуется создать систему видеоаналитики, входными данными для которой являются объекты (люди, роботы, самоходная техника, механизмы, станки и т. п.) и последовательности действий, которые они производят для достижения какого-то заранее известного результата. Как правило результатом является некоторая продукция, которая изготавливается или процесс, который контролируется. Правила, по которым действуют объекты описываются документами, называемыми технологические регламенты. Регламент можно соблюсти или не соблюсти в разной степени. Задача системы – предоставить инструменты и подсказать

алгоритмы для контроля выполнения регламента с момента, когда пользователь понял, что он хочет наблюдать и может сделать фото и видеосъемку наблюдаемого процесса, до момента генерации событий о факте выполнения или нарушения с метрикой, показывающей степень достижения результата для предоставления этой информации конечному пользователю.

Новаторским для систем видеоаналитики является признание факта, что в контексте технологических регламентов важное значение имеют не столько объекты, над которыми совершаются действия, не сами правильные действия, которые понятно как можно формализовать, а гарантией безопасности является формализация заведомо опасных действий. Суть предлагаемой системы заключается в том, чтобы обучить систему распознавать не наличие какого-то признака (например, детекции предмета на человеке), а аномалию среди уже известных ей правильных и неправильных действий.

Рассмотрим сначала систему как единой целое. Пусть заданы объекты-участники процесса, заданы и формализованы понятные пользователю действия над объектами (линейное перемещение, поворот, совпадение позиций объектов в пространстве), заданы и формализованы понятные пользователю результаты действий (регламенты). Так же заданы объекты-компоненты системы (программные модули, получающие на вход формальные данные и возвращающие формальный результат) и алгоритмы их взаимодействия (описаны протоколы коммуникации между программными модулями).

В нашей системе предполагается, что все взаимодействия являются машинно-машинными. В частности, человеко-машинное действие «добавить данные» является машинно-машинным между камерой смартфона объекта человек и файловой системой сервера, аналогично, человеко-машинное «оператор смотрит в монитор и видит событие» равносильно машинно-машинному система оповещения получила сообщение (отреагировала на событие) пайплайна мониторинга.

Важно отметить, что все процессы, несмотря на то, что они непрерывны во времени, детерминированы с вычислительной точки зрения, т. к. любое изменение в реальном времени происходит существенно дольше минимальной вычислительной единицы времени – тика процессора/контроллера. Примем, например, за единицу времени величину, обратную наименьшему общему кратному величине тика всех вычислительных устройств.

Все представленные выше сущности будут агентами в нашей мультиагентной среде видеоаналитики.

Назовем коммуникацией любое машинно-машинное инициативное обращение какого-то агента к одному или нескольким другим агентам либо получение обращения от кого-либо.

Посмотрим теперь на конкретного агента. Агенты согласно описанному выше бывают нескольких типов. Каждый тип агента знает, какими свойствами он обладает и какие данные он умеет обрабатывать. Также каждый агент знает, какова его локальная цель, и глобальную цель всей системы. Агент стремится достигнуть локальной цели с некоторой заданной вероятностью при условии выполнения глобальной цели системы с заданной пользователем вероятностью.

Агенты действуют в условиях шумов и неопределенности, т. е. каждый агент может не выполнить свою функцию с некоторой, только ему известной вероятностью.

В общем случае при необходимости достичь точного решения задача является NP-трудной, однако в реальных системах никогда не требуется и по факту не достигается значение 100%. Глобальная цель устанавливается соглашением об обслуживании и известна всем агентам, на параметры которого она влияет.

Поэтому построим модель на основе теории графов, похожую на конструкцию потоков в алгоритме Форда-Фолкерсона.

3. Математическая модель, программное моделирование

Задано множество агентов A . Пусть это универсальное множество. Рассмотрим некоторое подмножество множества агентов $S \subset A$. Назовём свойством s агента a принадлежность заданному подмножеству: $s(a) \in S$. Пусть у нас есть заданное семейство подмножеств. Предположим, что агент a_i обладает набором свойств $s_k(a_i)$.

Зададим отношение $R(a_i, a_j)$ между двумя агентами a_i и a_j . Назовем любое такое отношение коммуникацией.

По заданным отношениям R в момент времени t можно построить семейства графов $G(R, t)$, где вершинами будут агенты, а ребрами коммуникации. Каждая коммуникация имеет свою цену — вес ребра (числовой вектор). Каждый агент имеет свою локальную цель — вес вершины (числовой вектор вероятностей выполнить заданную цель, случайное число).

Пусть дано, что каждый такой граф ациклический ориентированный граф. Назовем вершины входящей степени 0 входами, а выходящей степени 0 -- выходами.

Назовём графом системы (глобальным графом) граф, вершинами которого являются все агенты, а ребрами все коммуникации по заданным отношениям. Назовём коммуникацию по заданному отношению локальной, а новую коммуникацию, имеющую цену как сумму всех цен локальных коммуникаций между заданной парой агентов, глобальной. Гарантируется, что такой граф имеет ровно один вход, ровно один выход и является ориентированным ациклическим в каждый момент времени.

Пусть для глобального графа задан ресурс — числовой вектор, состоящий из двух видов параметров: часть отвечает за вероятность прохождения вершины, часть — уменьшается на цену за прохождение ребра. Поместим весь ресурс в точку входа. Будем контролировать вероятности при прохождении вершин (достижение цели по заданному параметру) для оценки вероятности достижения выхода и учитывать ресурсы для оценки стоимости пути при условии его прохождения с вычисленной вероятностью. Считаем, что глобальная цель достигнута, когда на выходе вероятность попала в заданный доверительный интервал и существует хоть один путь, на который хватило ресурса, отвечающего за цену.

Основным результатом можно назвать тот факт, что данная модель позволяет построить программную систему, основанную на микросервисах, гарантированно решающую задачу с заданной вероятностью. Дело в том, что в системе мы можем использовать сколь угодно много лёгких агентов-детекторов различного назначения, за каждый из которых платим цену (время вычисления), однако после некоторого количества итераций для каждого конкретного вида набора входных данных (камеры монитора видят одни и те же группы регламентов) вычисляются наиболее дешёвые пути с наибольшей вероятностью, что является автоматической адаптацией нашей системы к пользовательским данным.

Программное моделирование данной системы осуществлялось на основе универсальной сети-трансформера MobileNet2 на маленьких данных различных типов на случайных графах. Были построены лёгкие модели детекторы, которые применялись в разработанном нами пайплайне управления низкого уровня (C-реализация на общей памяти с сетевым мониторингом на основе epoll). Пайплайн гарантирует ациклическость и направленность заданного графа исполнения. Основной определяющей ценой ребра выступало время, затрачиваемое на вычисления. Цель агентов-исполнителей — вероятность успешной детекции, классификации или кластеризации по заданному или обнаруженному признаку.

Онтология агентов (API) на основе json-файлов такова, что технически задача может быть решена иерархически. Т.е. сначала контролируемо обучаемая сеть находит подходящий кластер агентов-детекторов, а затем обрабатывает алгоритм принятия решения о способе классификации и снова происходит поиск очередного уточняющего кластера.

По результату тестирования сети из 14 агентов по видеоданным регламентных действий из интернета система смогла обучиться выдавать модель с допустимой точностью в 80% в 17 случаях из 21.

4. Заключение

Представленная работа является первой в цикле работ, посвященных созданию алгоритмов и концепции платформы видеоаналитики технологических регламентов для российской промышленности. Новизна данной работы заключается в применении совокупности идей и алгоритмов, разработанных и опробованных в принципиально разных предметных областях, в качестве агентов, решающих общую задачу – формализация объектов и действий, выделение их свойств, на первый взгляд никак не связанных друг с другом. Искусственный интеллект, который в отличие от человека не подвержен когнитивными искажениям, способен находить признаки сходства или различия формально. Т.е. для него неважно, был ли алгоритм разработан для поиска аномалий в электрической сети или предназначен для диагностики границ слоев в геологических ядрах. Для него важно, найдено ли отличие и, если применимо, отличие какого типа найдено. Это позволяет формализовать и автоматизировать поиск признака во входных данных агентом, ответственным за данный признак. Рандомизированный алгоритм идентификации параметров, работающий в условиях неопределенности, и позволяющий определить попадает ли модель системы в заданный класс с фиксированной вероятностью, успешно контролирует поведение агентов в кластере своего уровня, тем самым лишая систему высоконагруженного центра и делая ее полностью автономно распределенной.

Список литературы

1. Кутергин А.О. Алгоритм решения задачи распознавания действий объекта на видеопотоке // Вестник кибернетики. 2013. № 12. С. 111-117.
2. Азаренко К.А., Каунг Мьят Ньейн, Белов Ю.С. Обзор методов для распознавания действий человека // E-Scio. 2019. № 5 (32). <https://cyberleninka.ru/article/n/obzor-metodov-dlya-raspoznvaniya-deystviy-cheloveka> (дата обращения: 04.01.2024).
3. Liu Q., Han T., DARE: Dynamic Adaptive Mobile Augmented Reality with Edge Computing // 2018 IEEE 26th International Conference on Network Protocols (ICNP). Cambridge, UK, 2018. P. 1-11. doi: 10.1109/ICNP.2018.00011.
4. Xiaohui Liu, Chin-Seng Chua. Multi-agent activity recognition using observation decomposed hidden Markov models // Image and Vision Computing. 2006, Vol. 24, No. 2. P 166-175.
5. Brand M., Oliver N., Pentland A. Coupled hidden Markov models for complex action recognition // CVPR Conference, 1997.