

УДК 004.49

*А. А. Климов, Д. С. Лимонов, И. А. Топильский**

**ПОДХОД К АНАЛИЗУ ВИДЕОПОТОКА
С ЦЕЛЮ ВЫЯВЛЕНИЯ ФИЗИЧЕСКИХ ОБЪЕКТОВ
НА ОСНОВЕ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ**

Одной из наиболее востребованных и бурно развивающихся сфер ИТ-индустрии является группа технологий под общим названием *Big Data*. Эти технологии отличаются использованием обучающих программ для решения задач классификации, кластеризации, машинного зрения и многого другого. Предпосылками развития рассматриваемой отрасли стал экспоненциальный рост данных в современных распределенных сетевых информационных системах и развитие методов искусственного интеллекта, таких как искусственные нейронные сети, нечеткие семантические сети, генетические алгоритмы и другие [1].

Безусловно, катализатором активного развития *Big Data* стал тот факт, что результаты анализа больших объемов данных легко монетизируются. Так, невозможно себе представить современный маркетинг в рамках модели *b2c* без создания подробного портрета клиента, на основе анализа огромного количества заранее собранных о нем данных. На таких объемах обучающиеся алгоритмы существенно превосходят известные детерминированные подходы как с позиции вычислительной сложности, так по качеству построенного решения.

Наиболее распространенные подходы к созданию алгоритмов в парадигме *Big Data* связаны с предварительной настройкой его параметров на обучающей выборке. Например, самая распространенная концепция проектирования алгоритмов из рассматриваемой области – искусственные нейронные сети – предполагает предварительное обучение на некотором наборе входных данных с известным эталонным ответом. На каждом этапе сеть меняет свои внутренние параметры (веса синапсов), для того чтобы в среднем продуцировать на основе входных данных ответы более близкие к эталонным. Аналогичные задачи возникают и в других подходах парадигмы *Big Data*.

Обучающая выборка, кроме наличия эталонных ответов, должна обладать репрезентативностью. Это необходимо для того, чтобы искусственная нейронная сеть или алгоритм *Big Data* другого типа смогли обучиться с учетом всех скрытых параметров, влияющих на ответ. Данное требование накладывает ограничения не только на структуру выборки, но и на ее размер, так как огромный объем исследуемых данных

* Работа представлена в отборочном туре программы У.М.Н.И.К. 2016 г. в рамках Одиннадцатой межвузовской научной студенческой конференции Ассоциации «Объединенный университет им. В. И. Вернадского» «Проблемы техногенной безопасности и устойчивого развития» и выполнена под руководством канд. техн. наук, ст. преподавателя ФГБОУ ВО «ТГТУ» Д. В. Полякова.

порождает необходимость в сопоставимом размере обучающей репрезентативной выборки.

Представленные особенности *Big Data* позволяют сформулировать некоторые ограничения, которые данной группе технологий сложно преодолеть.

Во-первых, каждый алгоритм крайне узко специализирован. То есть, если обучить программу распознавать конкретные объекты (например, лица людей), она сможет решать данную задачу на уровне сопоставимом с человеческим. И, вместе с тем, будет совершенно бесполезна для распознавания других объектов.

Второе ограничение обусловлено наличием конечной обучающей выборки, на основе которой настраивает свои параметры алгоритм. Это означает, что «знания», полученные в ходе обучения алгоритмом, при дальнейшей его работе не уточняются и алгоритм совершенно не способен приспосабливаться к изменяющимся условиям.

Третьей проблемой, которая довольно часто возникает, является построение эталонных ответов для обучающей выборки. Допустим, исследуется задача распознавания такого сложного объекта, как автомобиль. Получить большое число фотографий, на которых изображены различные автомобили, снятые с разнообразных ракурсов не составляет сложности. Вместе с тем, задача разметки каждого пикселя, принадлежащего или не принадлежащего изображению автомобиля, становится крайне трудоемкой и практически невыполнимой задачей для формирования репрезентативной обучающей выборки, например, в размере свыше миллиона фотографий.

Вместе с тем, существует класс технологий, позволяющий выявлять объекты без предварительного обучения [2]. Один из таких подходов основан на анализе оптического потока (*optical flow*) видео с целью оценки движения (*motion estimation*). Выявление на разных кадрах точек, принадлежащих одним и тем же объектам, позволяет найти и создать некоторое представление реального физического объекта, появившегося на видео, без предварительного обучения алгоритма.

Вместе с тем, современные подходы ограничены проблемой апертуры, заключающейся в том, что отображение динамически изменяющегося, физического мира в последовательность кадров не является инъективным. То есть обратное преобразование – построение физических объектов по полученным кадрам – неоднозначно.

Основным подходом к решению данной задачи, на сегодняшний день, является регуляризация, которая, по сути, представляет собой использование некоторой априорной информации о выявляемом объекте. Вместе с тем, такой подход означает, что корректно восстановлены будут только объекты, о которых алгоритм «получил» априорное представление от разработчика, что вновь снижает широту возможностей подхода.

В нашей работе предлагается подход к преодолению проблемы графической апертуры, основанный на использовании генетических алгоритмов и гипотезе о том, что верное представление физического

объекта приведет к наиболее точному прогнозированию его (объекта) отображения на кадрах. Данная гипотеза позволяет предложить функцию полезности генетического алгоритма (*Fitness*), не связанную с природой объекта.

Пусть U – множество всех пикселей всех кадров видео. А T и F – множества верно и, соответственно, неверно спрогнозированных цветов пикселей для некоторой модели $m \in M$, где M – множество моделей. Отметим, что прогнозирование пикселя представляет собой прогнозирование его цвета, и модель m пытается спрогнозировать лишь часть пикселей, т.е. $T \cup F \neq U$. Определим *Fitness* модели как

$$Fitness = |T| - |F| - W_{Fitness}, \quad (1)$$

где $W_{Fitness}$ – «вес» самой модели, который должен коррелировать с размером данной модели в виртуальной памяти.

Благодаря учету «веса» самой модели, генетический алгоритм с функцией принадлежности вида (1) будет искать наиболее простой объект, правильно прогнозирующий цвет у максимального числа пикселей.

Рассмотрим еще вариацию *Fitness* при условии учета степени прогноза, т.е. пиксели с цветом, близким к заданному, будут увеличивать *Fitness* сильнее. Естественным подходом для развития модели в заявленном направлении является рассмотрение T и F как нечетких подмножеств U , формализуемых функциями принадлежности $\mu_T, \mu_F : U \rightarrow [0, 1]$. Значения этих функций в конкретном пикселе задают степень точности прогноза его цвета, сделанного моделью. Тогда, в силу того, что мощность множества обобщается кардинальным числом [3], выражение (1) принимает вид:

$$Fitness = \sum_{u \in U} (\mu_T(u) - \mu_F(u)) - W_{Fitness}. \quad (2)$$

В самом общем виде предлагаемый генетический алгоритм предполагает построение на первом шаге некоторой группы моделей физических объектов $M = \{m_1, m_2, \dots, m_n\}$. На втором шаге для каждой модели происходят операция нахождения множества пикселей, которые она прогнозирует, и вычисление функции полезности *Fitness*. Следующий шаг заключается в размножении, включающем в себя обмен моделей своими частями и мутации параметров. Вероятность размножения должна зависеть от расстояния между ближайшими пикселями спрогнозированной модели. В более строгой формулировке скрещиваться могут только те модели, у которых есть совпадающие или соседние прогнозируемые пиксели. Последующая селекция состоит из элиминации нежизнеспособных моделей ($Fitness < 0$) и вероятностного снижения численности популяции за счет рандомизированного исключения из нее моделей m с вероятностью P , вычисляющейся по формуле:

$$P = (Fitness_{\max} - Fitness_m) / (Fitness_{\max} - Fitness_{\min}), \quad (3)$$

где $Fitness_{\max}$ и $Fitness_{\min}$ – максимальное и минимальное значения функции полезности в популяции моделей соответственно.

Нормирование (3), кроме обеспечения выживания наилучшей и элиминации худшей моделей, позволяет катализировать конкуренцию за каждый пиксель в условиях высокого подобия в популяции. То есть важно не число пикселей, которые удалось спрогнозировать модели, а вес данной модели в сравнении с популяцией. Данный процесс повторяется, начиная со второго шага до достижения критерия останова, задаваемого оценкой качества прогноза лучшей модели.

Предлагаемый подход в перспективе позволит программам выявлять объекты постоянно, без предварительного обучения и, таким образом, снизит критичность сформулированных ранее ограничений *Big Data*. Результаты данной работы могут быть использованы как при решении проблем машинного зрения, анализа видеоданных, так для более тривиальных задач сжатого представления видеоинформации.

Список литературы

1. Громов, Ю. Ю. Выбор архимедовых норм и конорм на основе использования рядов Фурье / Ю. Ю. Громов и [др.] // Приборы и системы. Управление, контроль, диагностика. – 2012. – № 3. – С. 3–4.
2. Kaviani, H. R. Frame rate up-conversion using optical flow and patch-based reconstruction / H. R. Kaviani, S. S. Shirani // IEEE Transactions on Circuits and Systems for Video Technology. – 2015. – N 4. – P. 105–106.
3. Батыршин, И. З. Основные операции нечеткой логики / И. З. Батыршин. – Казань : Отечество, 2001. – 100 с.

*Кафедра «Информационные системы и защита информации»
ФГБОУ ВО «ТГТУ»*