

Соловьев Виктор Владимирович – Южный федеральный университет; e-mail: vvsolovev@sfedu.ru; г. Таганрог, Россия; тел.: +79043438844; кафедра систем автоматического управления; старший преподаватель; руководитель Центра проектной деятельности ИРТСУ ЮФУ.

Номерчук Александр Яковлевич – e-mail: aynomerchuk@sfedu.ru; тел. +79185945894; кафедра систем автоматического управления; старший преподаватель.

Soloviev Viktor Vladimirovich – Southern Federal University; e-mail: vvsolovev@sfedu.ru; Taganrog, Russia; phone: +79043438844; the department of automatic control systems; senior lecturer; head of the Center for Project Activities, Institute of Radio Engineering Systems and Control, SFU.

Nomerchuk Alexandr Yakovlevich – e-mail: aynomerchuk@sfedu.ru; phone: +79185945894; the department of automatic control systems; senior lecturer.

УДК 004.89

DOI 10.18522/2311-3103-2023-3-188-201

А.Г. Слепцов, И.С. Берешполов, Ю.А. Кравченко**РАЗРАБОТКА И ИССЛЕДОВАНИЕ МОДЕЛИ КЛАССИФИКАЦИИ
ВИДЕОИНФОРМАЦИИ***

Статья посвящена решению научной задачи классификации видеоконтента в условиях увеличения объемов информации. Компьютерное зрение является весьма актуальной областью применения технологий искусственного интеллекта для расширения возможностей различных поисковых и архивных систем. Авторами даны определения основным терминам исследуемой предметной области. Представлена формализованная постановка решаемой задачи. Приведена развернутая классификация возможных вариантов решения поставленной задачи. С быстрым развитием информационных технологий цифровой контент демонстрирует тенденцию к взрывному росту. Классификация спортивных видео имеет большое значение для архивирования цифрового контента на сервере. Многие алгоритмы интеллектуального анализа данных и машинного обучения достигли больших успехов во многих областях для применения (таких как классификация, регрессия и кластеризация). Однако большинство этих алгоритмов имеют общий недостаток, когда обучающая и тестовая выборки находятся в одном и том же пространстве признаков и подчиняются одному и тому же распределению. В этой статье рассматривается значимость решения задачи классификации видео и автоматического аннотирования содержания видеoinформации, а также разработана модель на основе глубокого обучения и больших данных. В рамках данного исследования авторами была разработана модель, которая повышает качество классификации видео, что позволяет улучшить результаты поиска. Результаты вычислительного эксперимента показывают, что предложенная модель может эффективно использоваться для распределения по классам видео событий в рамках спортивной предметной области на основе применения сверточной нейронной сети. При этом, обеспечивается высокая точность классификации видео спортивных тренировок. По сравнению с другими моделями предлагаемая имеет преимущества простой реализации, быстрой скорости обработки, высокой точности классификации а также высокой способности к обобщению.

Глубокое обучение; информационный поток; модель классификации; видеоконтент; компьютерное зрение.

* Исследование выполнено за счет гранта Российского научного фонда № 22-21-00316, <https://rscf.ru/project/22-21-00316/> в Южном федеральном университете.

A.G. Sleptsov, I.S. Bereshpolov, Yu.A. Kravchenko

DEVELOPMENT AND RESEARCH OF THE MODEL FOR VIDEO INFORMATION CLASSIFICATION

The article is devoted to solving the scientific problem of classifying video content in the face of an increase in the information volume. Computer vision is a very relevant field of artificial intelligence technologies application to expand the capabilities of various search and archive systems. The authors give definitions to the main terms of the studied subject area. A formalized statement of the problem to be solved is presented. A detailed classification of possible options for solving the problem is given. With the rapid development of information technology, digital content is showing an explosive growth trend. The classification of sports videos is of great importance for archiving digital content on the server. Many data mining and machine learning algorithms have made great strides in many application areas (such as classification, regression, and clustering). However, most of these algorithms have a common drawback when the training and test samples are in the same feature space and follow the same distribution. This article discusses the importance of solving the problem of the video information content classification and automatic annotation, and also develops a model based on deep learning and big data. As part of this study, the authors developed a model that improves the quality of video classification, which improves search results. The results of the computational experiment show that the proposed model can be effectively used to classify video events within the sports subject area based on the use of a convolutional neural network. At the same time, high accuracy of sports training video classification is provided. Compared with other models, the proposed model has the advantages of simple implementation, fast processing speed, high classification accuracy, and high generalization ability.

Deep learning; information flow; classification model; video content; computer vision.

Введение. В условиях быстрого развития компьютеров, сетей, мультимедиа и других информационных технологий мультимедийные данные демонстрируют тенденцию к экспоненциальному росту. Видео содержит информацию со сложной структурой и большим объемом данных. Для людей с ослабленным зрением автоматическое описание видео в сочетании с технологией преобразования текста в речь переводит текст в непрерывный естественный язык для общения. Это помогает лучше понять содержание видеоинформации, тем самым облегчая жизнь слабовидящим. В области исследования автоматического аннотирования видео, интеллектуальный анализ и понимание видеоинформации, основанные на моделировании действий человека, в последние годы стали популярной исследовательской проблемой в области компьютерного зрения и распознавания образов [1]. Компьютерное зрение имеет широкую перспективу применения в области интеллектуальной помощи жизни человека, расширенного взаимодействия его с компьютером, а также поиска видео на основе анализа контента.

Растущие и увеличивающиеся объемы и вычислительная сложность данных делают недостаточными возможности определенных алгоритмов и мощности одного компьютера. Использование распределенных вычислительных платформ, чтобы справиться с огромным потреблением временной и пространственной сложности алгоритмов в средах больших данных, а также для решения проблемы концептуального дрейфа в потоках данных, стало обязательной необходимостью.

С точки зрения теоретического исследования, изучение автоматического аннотирования видеоинформации является междисциплинарной темой, которая объединяет несколько дисциплин, таких как анализ больших данных, машинное обучение, распознавание образов, компьютерное зрение и когнитивные технологии, что обеспечивает необходимую методологическую базу для этих научных областей, и ее углубленное изучение будет способствовать развитию смежных областей. С развитием нейронных сетей и появлением крупномасштабных наборов данных в области классификации изображений и распознавания образов в ряде подходов

были предприняты попытки использования сверточных нейронных сетей для изучения семантики представления изображений, а затем рекуррентных нейронных сетей для достижения соответствия получаемых решений естественному языку.

Традиционное контролируемое обучение в основном представляет собой обучение с одной меткой, в то время, как реальные целевые образцы часто бывают сложными, имеют несколько семантик и содержат несколько меток. В последние годы автоматическое понимание видеоинформации постепенно становится популярным направлением исследований в области компьютерного зрения. По сравнению с задачей распознавания изображений содержание видео включает больше информации, и одна метка не может полностью охарактеризовать видеоконтент, поэтому большинство задач для автоматического понимания видеоконтента требует введения ряда специальных меток. Эффективный анализ корреляций между метками является ключом к успеху обучения с несколькими такими маркерами. По силе корреляций интеллектуального анализа данных алгоритмы с несколькими метками можно разделить на стратегии следующих трех типов:

1. Стратегия первого типа: игнорирование корреляции между метками, например, разложение нескольких меток на несколько независимых задач бинарной классификации;

2. Стратегия второго типа: рассмотрение парных корреляций между метками, такие как ранжирование релевантных и нерелевантных меток;

3. Стратегия третьего типа: рассмотрение корреляции между несколькими метками, например, учитывая влияние всех других меток на каждую метку.

Еще один подход основан на алгоритмически применимых методах. Методы, применимые к алгоритму – это методы, расширяющие возможности конкретного алгоритма для обработки данных с несколькими метками.

1. Постановка задачи классификации видеоинформации. Огромный объем ресурсов данных, таких как поведенческие характеристики пользователей, необходимо дополнительно анализировать и добывать для создания коммерческих моделей классификации видео, чтобы предоставлять их производителям и продавцам для улучшения качества продукта [2]. Технологии больших данных и глубокого обучения не только помогут увеличить добавленную стоимость продукта, но и максимизируют ценность клиентского опыта.

Информационный поток – это набор последовательностей данных. В общем, информационный поток данных можно рассматривать как динамическую коллекцию данных, которая неограниченно увеличивается с течением времени. Он используется в области сетевого мониторинга, сенсорной сети, аэрокосмической отрасли, метеорологических измерений и контроля, финансовых услуг и т.п. Изучение информационных потоковых данных представляет собой исследование нового типа модели обработки данных и, следовательно, требует иного подхода к интеллектуальному анализу запросов данных, по сравнению с традиционными методами. Данный подход должен быть адаптирован для реализации сценариев обработки потоковых данных на основе запросов и распределенных вычислений. Это требует построения комплекса алгоритмов для максимально быстрой обработки запроса и извлечения ценных данных, в рамках решения практических задач потоковой обработки и передачи данных в информационной среде. Исследования по анализу и обработке больших объемов данных в основном проводятся в двух направлениях: запрос информационных потоковых данных и интеллектуальный анализ информационных потоковых данных [3–4].

Для решения задачи классификации видеоинформации предварительно необходимо обработать входные данные, представленные в виде видеороликов (в данном исследовании рассмотрим предметную область с обработкой видеоконтента

спортивных мероприятий). Съемка видео на спортивном тренировочном объекте обычно делится на дистанционное видео, видео среднего расстояния и видео близкого расстояния. Для более точной классификации видео спортивных тренировок входные видеоролики разбиваются на сегменты равной продолжительности, что приводит к получению субвидеосегментов:

$$V = \{V_1, V_2, V_3, \dots, V_M\},$$

$$v_j = \{v_{j1}, v_{j2}, \dots, v_{jq}, \dots, v_{jm}\},$$

где V представляет сегмент видео, соответствующий конкретному спортивному событию. В формуле v_j представляет видеоизображение кадра j , $j = 1, 2, 3, \dots, N$, а N задает количество кадров, преобразованных в изображение видеокдра входного видеосегмента.

В приведенном выше уравнении $jm = pm, j \neq p, j, p = 1, 2, \dots, M$ $q = 1, 2, \dots, m$ v_j представляет сегмент j -го суб-видео после сегментации видео, v_{jq} представляет изображение q -го кадра в j -м сегменте суб-видео и M представляет количество сегментов суб-видео. Временной интервал сегментированного видеополя оказывает определенное влияние на результаты классификации.

Опишем процедуру извлечения поля вектора движения. Пусть размер видео спортивной тренировки будет $M \times N \times T$, где $M \times N$ обозначают разрешение, а T обозначает длину видеоряда. Видео разделено на блоки $K \times L$, а $h \times v$ – размер блока, где $h = M/K$. Примем за основу прямоугольную систему координат для отображения вектора движения. Схема отображения векторного поля движения прямоугольной системы координат показана на рис. 1.

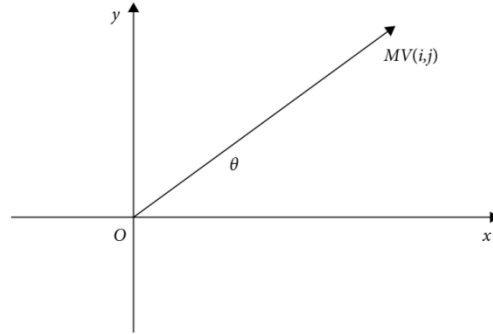


Рис. 1. Схема отображения поля вектора движения

На рис. 1 $MV(i, j)$ – это блок с позицией (i, j) , а $\theta \in [0, 2\pi)$ – это направление вектора движения C . C_x – составляющая вектора движения i -го блока в горизонтальном направлении (x), а C_y – составляющая вектора движения i -го блока в вертикальном направлении (y), ρ – интенсивность движения блока, тогда:

$$\begin{cases} \rho = \sqrt{C_x^2 + C_y^2} \\ \sin(\theta) = \frac{C_y}{\rho} \\ \tan(\theta) = \frac{C_y}{C_x} \end{cases}$$

Система координат непрерывных видеок кадров располагается в хронологическом порядке Q и разбивается на равные угловые сектора по положительному направлению x , p квантуется на R интервалы, а затем строятся гистограммы p и θ , так можно получить:

$$\begin{cases} Hist_q = \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^C q_i^t, & q \in [1, Q] \\ Hist_r = \frac{1}{q} \sum_{t=1}^T \sum_{i=1}^C r_i^t, & r \in [1, R] \end{cases}.$$

В данной системе q_i^t представляет количество векторов движения в квадранте q в кадре t и r_i^t представляет количество p квантованных r в кадре t [5].

Математическое ожидание и дисперсия вектора движения в направлениях x и y используются для оценки движения в блоке, а именно:

$$\begin{cases} \mu_x = \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^C C_{x,i}^t \\ \mu_y = \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^C C_{y,i}^t \\ \sigma_x^2 = \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^C (C_{x,i}^t - \mu_x)^2 \\ \sigma_y^2 = \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^C (C_{y,i}^t - \mu_y)^2 \end{cases}.$$

В данной системе $C_{x,i}^t$ и $C_{y,i}^t$ представляют компоненты вектора движения i макроблока в направлениях x и y в кадре, а μ_x, μ_y, σ_x^2 и σ_y^2 представляют математическое ожидание и дисперсию вектора движения макроблока в направлениях x и y соответственно.

2. Разработка модели классификации видеoinформации на основе сверточной нейронной сети. Сверточная нейронная сеть обычно состоит из нескольких слоев свертки, слоев понижающей выборки и слоев нормализации. Наконец, двумерная карта объектов соединяется с вектором и вводится в окончательный классификатор через полностью связанный слой, чтобы получить значение вероятности.

В сверточном слое функции верхнего слоя свертываются с помощью обучаемого ядра свертки, а затем выходные функции могут быть получены с помощью функции активации. Каждый выход может быть объединен для свертки значений нескольких входов [6]:

$$x_j^1 = f \left(\sum_{i \in M_j} x_i^{l=1} \cdot k_{ij}^{l-1} + b_j^1 \right).$$

В приведенном выше уравнении M_j представляет собой набор входных функций, связанных ядром свертки M_j и определяет связь между ядром свертки и входным слоем. Карта выходных объектов получается путем свертки ядра входной карты объектов. Предполагая, что каждое ядро свертки извлекает шаблон, любая выходная карта признаков соответствует признаку, а каждое ядро свертки эквивалентно карте признаков. Это связано с тем, что слой свертки использует технологию распределения веса, то есть каждый нейрон использует один и тот же вход проверки свертки для выполнения свертки, и каждый нейрон связан только с некоторыми входными нейронами, что уменьшает количество параметров слоя свертки. Функция f – это функция активации нейронов, которая обычно является нелинейной функцией.

Входом слоя свертки является несколько двумерных плоскостей, и каждое ядро свертки связано со всеми входными каналами. Свертка выполняется в трехмерном пространстве для получения выходных данных отклика положения. Нако-

нец, алгоритм проверяет свертку всего входного пространства, чтобы получить карту объектов. Обычно в каждом слое свертки устанавливается несколько ядер свертки, и каждое ядро свертки извлекает разные признаки, так, что каждая карта признаков представляет плоскость признаков, извлекаемую соответствующим ядром свертки.

Целью слоя понижающей выборки является повышение устойчивости сети к небольшой деформации входных выборок, чтобы улучшить производительность сети при обобщении. Где y_{ijk} используется для представления выходных данных нейрона в нижнем слое дискретизации [7]. Уровень понижающей дискретизации может быть представлен следующим выражением:

$$y_{ijk} = \sum_{pq} w_{pq} x_{i,j+p,k+q},$$

где w_{pq} – нормализованное взвешенное окно, которое может выполнять выборку каждой входной карты объектов без пересечения разных карт объектов. Количество выходных карт объектов в слое понижающей выборки такое же, как и количество входных карт объектов, что снижает разрешение каждой карты объектов.

Слой нормализации очень важен для повышения производительности нейронной сети. В модели сверточной нейронной сети уровень нормализации включает в себя нормализацию вектора признаков одной и той же карты признаков и карты признаков, расположенной в разных картах признаков, что усиливает карту признаков с более высоким значением отклика и управляет различными ядрами свертки для изучения разных шаблонов. Операция вычитания и нормализации в заданном местоположении на самом деле представляет собой значение местоположения за вычетом взвешенного значения каждого пикселя в окрестности. Вес может быть определен взвешенным по Гауссу окном. Нормализация деления является общим алгоритмом нормализации, который может усилить разницу значений отклика и улучшить эффект высокой характеристики значения отклика.

Нормализация локального отклика является распространенным алгоритмом нормализации в сверточных сетях. Значение отклика может быть задано следующим выражением:

$$b_{x,y}^i = \frac{a_{x,y}^i}{(K + \alpha \sum_{j=\max(0, i-\frac{n}{2})}^{\min(N-1, i+\frac{n}{2})} (a_{x,y}^j)^2)^\beta},$$

где $a_{x,y}^i$ представляет собой значение i -й карты входных объектов в координате (x, y) . N представляет количество входных карт объектов, а n является нормализацией на соседних картах n [8].

Слой нормализации локального отклика содержит три регулируемых параметра, а именно количество карт признаков n и параметров α и β . Все слои нормализации используют одну и ту же настройку параметров $n = 5$, $\alpha = 0,0005$, $\beta = 0,5$.

Полносвязный слой обычно находится в верхней части нейронной сети, которая образует традиционную многослойную перцептивную сеть вместе со слоем принятия решений для классификации признаков, извлеченных из слоя свертки. Переоснащение сверточной нейронной сети в основном вызвано большим количеством параметров в полносвязном слое. Технология Dropout обычно добавляется к полносвязному слою, и некоторые нейроны выбираются случайным образом для участия в обучении, чтобы предотвратить переоснащение сети.

Многослойная сверточная нейронная сеть состоит из вышеуказанных пяти слоев нейронов, которые выполняют разные функции и должны быть объединены по определенным правилам для достижения лучших результатов. Среди пяти слоев нейронов только слой свертки и полностью связанный слой содержат обучае-

мые параметры. Слой свертки так же может сохранять входную информацию о пространственном положении, которая требуется для слоя выборки. Слой свертки обычно используется попеременно со слоем понижающей дискретизации, так что разные слои свертки могут извлекать различные характеристики. Полностью связанный слой уничтожит информацию о положении характерных плоскостей и разницу между каждой характерной плоскостью. Полносвязный слой обычно используется как часть конечного многоуровневого перцептивного классификатора, который объединяет слой свертки и слой понижающей дискретизации для извлечения признаков и отправки их на уровень принятия решений для классификации.

На основе выходных данных сверточной нейронной сети события тестовой и эталонной видеоинформации сравниваются методом сопоставления событий. Учитывая символы наблюдения класса видео, модель сверточной нейронной сети с прохождением нескольких состояний обучается с использованием признаков, извлеченных из загруженных видеокладов для получения последовательности событий (вероятность события и соответствующий переход состояния) в соответствующем эталонном видео. Эталонная последовательность событий используется для создания словаря исследуемого тематического события. Для события с определенным переходом состояния (k, l) в эталонном событии распределение вероятности события аппроксимируется функцией плотности Гаусса $N(\mu_{kl}, \sigma_{kl})$, где μ_{kl} и σ_{kl} представляют собой среднее значение и дисперсию функции плотности соответственно, что определяется следующими выражениями:

$$\mu_{kl} = \frac{1}{L_1} \sum_{t=1}^{L_1} e_t^p(k, l),$$

$$\sigma_{kl} = \sqrt{\frac{1}{L_1} \sum_{t=1}^{L_1} (e_t^p(k, l) - \mu_{kl})^2}.$$

Каждому переходу состояния присваивается среднее значение и дисперсия для представления вероятности $e_t^p(k, l)$ события, происходящего в категории. Для видеоклипов спортивных тренировок, которые не появляются на этапе обучения, получение событий происходит на основе применения эталонной модели нейронной сети свертки. Тогда $e_t^p(k, l)$ обозначает вероятность события перехода состояния (k, l) во время t , когда тестовая последовательность в символе наблюдения обеспечивает эталонную модель. Обозначим через L_2 число символов наблюдения в тестовой последовательности. Сходство между тестовым видеоклипком и эталонной моделью выражается следующим уравнением [9]:

$$s = \frac{1}{L_2} \sum_{t=1}^{L_2} \frac{1}{\sqrt{2\pi}\sigma_{kl}} \exp \left[-\frac{(e_t^p(k, l) - \mu_{kl})^2}{2\sigma_{kl}^2} \right].$$

Сравнивается значение сходства s между видеороликами и всеми видами изучаемых спортивных событий, после чего они классифицируются в категорию с наибольшим значением порога сходства.

3. Разработка высокоуровневого алгоритма классификации видеоинформации. При наличии базовой структуры основной проблемой, которую необходимо решить в ходе исследования, является отсутствие экспериментальных данных. В рамках представленного исследования будем использовать общедоступный набор данных спортивных видео Sport-1M, который в настоящее время является крупнейшим набором данных классификации видеоинформации, состоящим из 1,1 миллиона спортивных видео. Каждое видео относится к одной из 487 спортивных категорий, и в этом наборе данных не проиндексированы разложенные действия определенной категории видеоинформации [10]. В настоящее время не хватает тестовых наборов данных, которые позволяют получать описания бесплатных видеофрагментов.

Для расчета дискриминационной мощности видеокадров предложен ряд методов решения этой проблемы. Одной из часто используемых стратегий является использование механизма внимания. Механизм внимания встраивается в существующую сеть описания видео для вычисления весов между различными видеокадрами и повышения точности автоматического аннотирования видеoinформации. Как показано на рис. 2, во-первых, создается набор данных о декомпозиции, для видеоданных извлечение признаков выполняется сверточной нейронной сетью, а для размеченного текста строится словарь для извлечения соответствующих признаков словаря [11]. Рис. 2 показывает структуру, в которой тренировочный набор вводится в модель обучения для преобразования видеoinформации в текст. Полученный текст достаточно точно отражает смысл соответствующего видеоконтента.

Входные данные для решаемой задачи включают в себя видеопоследовательности и текстовые последовательности. Параметры видео сначала извлекаются сверточной нейронной сетью, а затем реализуется обработка текста на естественном языке.

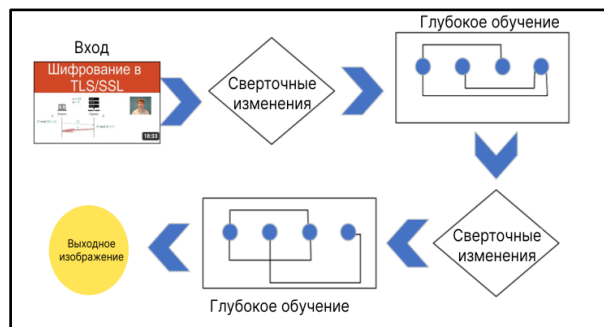


Рис. 2. Структура алгоритма классификации видеoinформации

Сверточные нейронные сети обладают некоторой степенью инвариантности к геометрическим преобразованиям, деформациям и освещению [12]. Обученные сверточные нейронные сети могут сканировать все изображение с небольшими вычислительными затратами и, следовательно, широко используются для извлечения признаков изображения. А при использовании выделения признаков особая морфология признаков не учитывается.

В данном исследовании применяется векторное представление для преобразования дескрипторов видеороликов. Слова в аннотированном тексте спортивных видеороликов сначала подсчитываются, чтобы составить словарь. Вычислительный процесс включает в себя ввод – это предложение и вывод – функция [13]. Расчет выполняется путем подсчета общего количества всех слов, описанных в наборе данных, а затем представления каждого слова в виде вектора. Схема данного процесса обработки видеопотока показана на рис. 3.

Для проверки достоверности и эффективности представленных модели и алгоритма классификации видеoinформации в качестве тестовых выборок были взяты восемь наборов данных, которые часто используются при классификации видеoinформации в исследуемой предметной области. Наборы данных включают восемь типов спортивных обучающих видеороликов, таких как баскетбол, волейбол и футбол. В качестве моделей сравнения выбраны модель опорных векторов (SVM) и модель HMM.

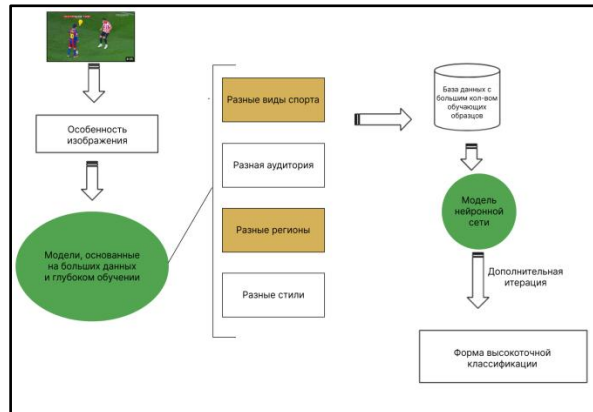


Рис. 3. Схема процесса обработки видеопотока

4. Результаты вычислительного эксперимента. Для определения точности классификации были выбраны следующие метрики: скорость отзыва (recall ratio) и скорость точности (precision ratio), эти метрики выбраны в качестве важных показателей для оценки эффективности классификации на основе применения предложенной модели. Здесь n_c используется для представления количества правильных результатов распознавания, n_m используется для представления количества неправильных результатов распознавания и n_f используется для представления количества неудачных результатов распознавания [14–21]. Для оценки выбирается среднее значение пяти экспериментов и устанавливается соотношение 2:1 для случайного разделения обучающей выборки и тестовой выборки. Уравнение индекса оценки выглядит следующим образом:

$$\text{recall ratio} = \frac{n_c}{(n_c + n_m)},$$

$$\text{precision ratio} = \frac{n_c}{(n_c + n_f)}.$$

Статистика результатов сравнения точности различных наборов данных и различных видов классификации видеoinформации представлена на рис. 4

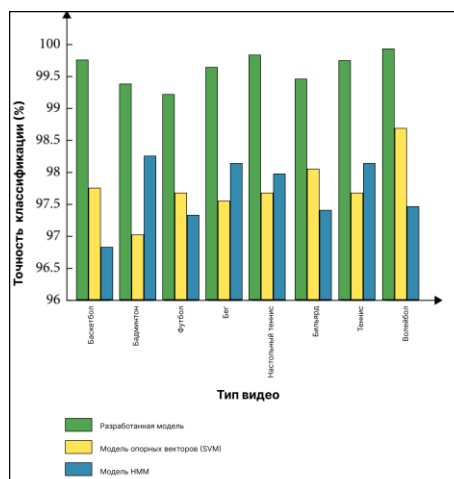


Рис. 4. Сравнение точности классификации видеoinформации

Как видно из результатов вычислительного эксперимента на рис. 4, при разных наборах данных и разных видах мероприятий точность отзыва классификации видеоинформации, классифицированной по предлагаемой модели, выше 99%, а у двух других моделей данный показатель значительно ниже, что подтверждает тот факт, что разработанная модель имеет более высокую точность классификации видеоинформации.

Статистика результатов сравнения частоты повторного просмотра различных наборов данных и различных видов классификации видеоинформации представлена на рис. 5.

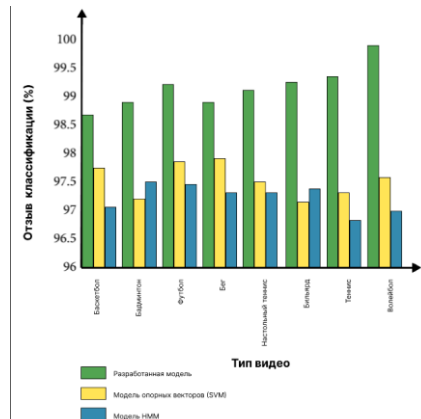


Рис. 5. Сравнение показателей классификации видеоинформации

Как видно из результатов вычислительного эксперимента на рис. 5, при разных наборах данных и разных видах спортивной подготовки показатель повторного просмотра тренировочных видеороликов, классифицированных по предложенной модели, выше 98,5%, а показатель повторного просмотра тренировочных видеороликов, классифицированных по другим моделям значительно ниже, что подтверждает эффективность предложенной модели.

Видеоданные состоят из более чем одного разложенного действия, поэтому каждый элемент видеопотока индексируется как минимум одной категорией, а категории кодируются положительными целыми числами (начиная с 1), всего 31 категория. Каждая категория может потребовать более одного слова для своего описания. Половина категорий встречается очень редко (меньше 10 раз), при этом относительно небольшое количество категорий встречается особенно часто. Рис. 6 показывает анализ частоты 16 категорий, которые встречаются более 10 раз, причем только несколько категорий имеют особенно большое количество появлений.

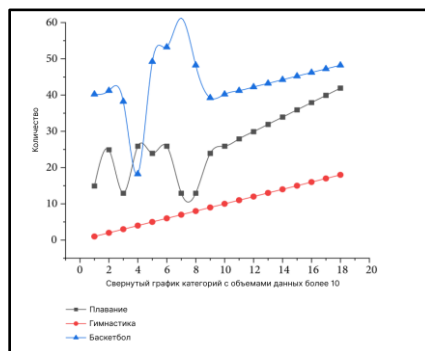


Рис. 6. Зависимость частоты появления категорий с объемом данных более 10

Приведенные выше результаты вычислительного эксперимента показывают, что предложенная модель достаточно точно классифицирует видеoinформацию, что подтверждает ее эффективность. Данная модель использует метод глубокого обучения для решения задачи классификации. Временная сложность предложенного алгоритма классификации в представленном примере составила $O(n^2)$, где n – количество анализируемых алгоритмом входных данных.

Заключение. В данной статье представлена разработка модели, которая повышает качество классификации видео, что позволяет улучшить результаты поиска. Результаты вычислительного эксперимента показали, что предложенная модель может эффективно использоваться для распределения по классам видео событий в рамках спортивной предметной области на основе применения сверточной нейронной сети.

Основными результатами проведенного исследования являются следующие:

1. Авторами даны определения основным терминам исследуемой предметной области. Представлена формализованная постановка решаемой задачи. Приведена развернутая классификация возможных вариантов решения поставленной задачи;

2. Приведена развернутая классификация возможных вариантов решения поставленной задачи;

3. Построена модель классификации спортивной видеoinформации, основанная на методе глубокого обучения. Сверточная нейронная сеть с глубоким обучением используется для целей классификации в предлагаемом исследовании. После классификации выполняется операция сопоставления событий, и осуществляется классификация видеoinформации по сходству;

4. Разработан высокоуровневый алгоритм классификации видеoinформации, реализующий основные преимущества построенной модели.

Для оценки эффективности предложенных решений разработано программное приложение и проведен вычислительный эксперимент. Полученные результаты проведенных экспериментальных исследований подтверждают эффективность предложенной модели и алгоритма классификации спортивной видеoinформации. Временная сложность представленного алгоритма является полиномиальной. При выполнении исследования решены все поставленные задачи, цель данной работы достигнута.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Черкасова И.С. Классификация видео контента на основе сверточных нейронных сетей // E-Scio. – 2021. – № 12 (63). – С. 395-405.
2. Киреев В.С., Федоренко В.И. Использование методов векторизации текстов на естественном языке для повышения качества контентных рекомендаций фильмов // Современные наукоемкие технологии. – 2018. – № 3. – С. 102-106.
3. Никитин И.К. Обзор методов комплексного ассоциативного поиска видео // Вестник Новосибирского государственного университета. Серия: Информационные технологии. – 2014. – Т. 12, № 4. – С. 71-82.
4. Кузнецов А.А. Обзор возможности обучения роботов при помощи алгоритмов глубокого обучения // Вестник науки. – 2022. – Т. 2, № 6(51). – С. 239-243.
5. Никитин И.К. Элементы ассоциативного поиска по видео // Новое слово в науке: перспективы развития. – 2014. – № 1 (1). – С. 223-230.
6. Полещук Х.А., Соломатин Д.И. Классификация видео по описаниям // Сб. студенческих научных работ факультета компьютерных наук ВГУ: Сб. статей. В 2-х ч. Ч. 2 / под ред. Д.Н. Борисова. – Вып. 13. – Воронеж: Воронежский государственный университет, 2019. – С. 183-189.
7. Бова В.В., Кравченко Ю.А., Кулиев Э.В., Курейчик В.В. Моделирование поведения субъекта в Интернет-сервисах на основе модифицированного алгоритма бактериальной оптимизации // Информационные технологии. – 2019. – Т. 25, № 7. – С. 397-404.

8. Братских И.А. Искусственный интеллект, машинное обучение, глубокое обучение и нейронные сети // Правопорядок в России: проблемы совершенствования: Сб. материалов XIV Всероссийской конференции, Москва, 11–14 февраля 2020 года. – М.: Московский университет МВД России им. В.Я. Кикотя, 2020. – С. 285-286.
9. Городничев Д.Ю. Машинное обучение и глубокое обучение // Современные проблемы лингвистики и методики преподавания русского языка в ВУЗе и школе. – 2022. – № 38. – С. 278-281.
10. Hu F., Hao Q., and Bao K. A survey on software-defined network and openflow: from concept to implementation // IEEE Communications Surveys & Tutorials. – 2014. – Vol. 16, No. 4. – P. 2181-2206.
11. Naseer S., Saleem Y., Khalid S. et al. Enhanced network anomaly detection based on deep neural networks // IEEE Access. – 2018. – Vol. 6. – P. 48231-48246.
12. Vo A.H., Hoang Son L., Vo M.T., and Le T. A novel framework for trash classification using deep transfer learning // IEEE Access. – 2019. – Vol. 7. – P. 178631-178639.
13. Tong H., Chu J., and Shen J. Image classification technology based on convolutional neural network // Science and Technology Vision. – 2017. – Vol. 1, No. 032. – P. 36-37.
14. Zhang F. and Yan J. Cloud image classification method based on deep convolutional neural network / Xibei Gongye Daxue Xuebao // Journal of Northwestern Polytechnical University. – 2020. – Vol. 38, No. 4. – P. 740-746.
15. Hecker A. Knowledge Beyond the Individual. Making Sense of a Notion of Collective Knowledge in Organization Theory // Organization Studies. – 2012. – Vol. 33. – P. 423-445.
16. Bova V.V., Kureichik V.V., Nuzhnov E.V. The Combined Method of Semantic Similarity Estimation of Problem Oriented Knowledge on the Basis of Evolutionary Procedures // Artificial Intelligence Trends in Intelligent Systems. – 2015. – P. 74-83.
17. Rodzin S., Rodzina L. Theory of Bioinspired Search for Optimal Solutions and its Application for the Processing of Problem-Oriented Knowledge // 8th IEEE International Conference on Application of Information and Communication Technologies. – 2014. – P. 142-147.
18. Dolinina O., Suchkova N. Formal Models of the Structural Errors in the Knowledge Bases of Intellectual Decision Making Systems // Artificial Intelligence Trends in Intelligent Systems. – 2015. – P. 156-167.
19. Jensen J. A Systematic Literature Review of the Use of Semantic Web Technologies in Formal Education // Br. J. Edu. Technol. – 2019. – Vol. 50. – P. 505-517.
20. Гулякина Н.А., Давыденко И.Т. Семантические модели и метод согласованной разработки баз знаний // Программные продукты и системы. – 2020. – № 3. – С. 420-429.
21. Шалимов П.Ю. Модели количественной шкалы оценки семантической информации // Автоматизация и моделирование в проектировании и управлении. – 2020. – № 2 (8). – С. 24-32.

REFERENCES

1. Cherkasova I.S. Klassifikatsiya video kontenta na osnove svertochnykh neyronnykh setey [Video content classification based on convolutional neural networks], *E-Scio*, 2021, No. 12 (63), pp. 395-405.
2. Kireev V.S., Fedorenko V.I. Ispol'zovanie metodov vektorizatsii tekstov na estestvennom yazyke dlya povysheniya kachestva kontentnykh rekomendatsiy fil'mov [Using methods of vectorization of texts in natural language to improve the quality of movie content recommendations], *Sovremennye naukoemkie tekhnologii* [Modern high technologies], 2018, No. 3, pp. 102-106.
3. Nikitin I.K. Obzor metodov kompleksnogo assotsiativnogo poiska video [Overview of methods for complex associative video search], *Vestnik Novosibirskogo gosudarstvennogo universiteta. Seriya: Informatsionnye tekhnologii* [Bulletin of the Novosibirsk State University. Series: Information technologies], 2014, Vol. 12, No. 4, pp. 71-82.
4. Kuznetsov A.A. Obzor vozmozhnosti obucheniya robotov pri pomoshchi algoritmov glubokogo obucheniya [Overview of the possibility of training robots using deep learning algorithms], *Vestnik nauki* [Bulletin of science], 2022, Vol. 2, No. 6(51), pp. 239-243.
5. Nikitin I.K. Elementy assotsiativnogo poiska po video [Elements of associative video search], *Novoe slovo v nauke: perspektivy razvitiya* [New word in science: development prospects], 2014, No. 1 (1), pp. 223-230.

6. *Poleshchuk Kh.A., Solomatin D.I.* Klassifikatsiya video po opisaniyam [Classification of videos by descriptions], *Sb. studenteskikh nauchnykh rabot fakul'teta komp'yuternykh nauk VGU: Sb. statey* [Collection of student scientific papers of the Faculty of Computer Science of the Voronezh State University: Collection of articles]. In 2nd part. Part 2, ed. by D.N. Borisova, Issue 13. Voronezh: Voronezhskiy gosudarstvennyy universitet, 2019, pp. 183-189.
7. *Bova V.V., Kravchenko Yu.A., Kuliev E.V., Kureychik V.V.* Modelirovanie povedeniya sub'ekta v Internet-servisakh na osnove modifitsirovannogo algoritma bakterial'noy optimizatsii [Modeling the subject's behavior in Internet services based on a modified bacterial optimization algorithm], *Informatsionnye tekhnologii* [Information technologies], 2019, Vol. 25, No. 7, pp. 397-404.
8. *Bratskikh I.A.* Iskusstvennyy intellekt, mashinnoe obucheniye, glubokoe obucheniye i neyronnye seti [Artificial intelligence, machine learning, deep learning and neural networks], *Pravoporyadok v Rossii: problemy sovershenstvovaniya: Sb. materialov XIV Vserossiyskoy konferentsii, Moskva, 11–14 fevralya 2020 goda* [Law and order in Russia: problems of improvement: Proceedings of the XIV All-Russian Conference, Moscow, February 11–14, 2020]. Moscow: Moskovskiy universitet MVD Rossii imeni V.Ya. Kikotya, 2020, pp. 285-286.
9. *Gorodnichev D.Yu.* Mashinnoe obucheniye i glubokoe obucheniye [Machine learning and deep learning], *Sovremennyye problemy lingvistiki i metodiki prepodavaniya russkogo yazyka v VUZe i shkole* [Modern problems of linguistics and methods of teaching the Russian language at the university and school], 2022, No. 38, pp. 278-281.
10. *Hu F., Hao Q., and Bao K.* A survey on software-defined network and openflow: from concept to implementation, *IEEE Communications Surveys & Tutorials*, 2014, Vol. 16, No. 4, pp. 2181-2206.
11. *Naseer S., Saleem Y., Khalid S. et al.* Enhanced network anomaly detection based on deep neural networks, *IEEE Access*, 2018, Vol. 6, pp. 48231-48246.
12. *Vo A.H., Hoang Son L., Vo M.T., and Le T.* A novel framework for trash classification using deep transfer learning, *IEEE Access*, 2019, Vol. 7, pp. 178631-178639.
13. *Tong H., Chu J., and Shen J.* Image classification technology based on convolutional neural network, *Science and Technology Vision*, 2017, Vol. 1, No. 032, pp. 36-37.
14. *Zhang F. and Yan J.* Cloud image classification method based on deep convolutional neural network, *Xibei Gongye Daxue Xuebao, Journal of Northwestern Polytechnical University*, 2020, Vol. 38, No. 4, pp. 740-746.
15. *Hecker A.* Knowledge Beyond the Individual. Making Sense of a Notion of Collective Knowledge in Organization Theory, *Organization Studies*, 2012, Vol. 33, pp. 423-445.
16. *Bova V.V., Kureichik V.V., Nuzhnov E.V.* The Combined Method of Semantic Similarity Estimation of Problem Oriented Knowledge on the Basis of Evolutionary Procedures, *Artificial Intelligence Trends in Intelligent Systems*, 2015, pp. 74-83.
17. *Rodzin S., Rodzina L.* Theory of Bioinspired Search for Optimal Solutions and its Application for the Processing of Problem-Oriented Knowledge, *8th IEEE International Conference on Application of Information and Communication Technologies*, 2014, pp. 142-147.
18. *Dolinina O., Suchkova N.* Formal Models of the Structural Errors in the Knowledge Bases of Intellectual Decision Making Systems, *Artificial Intelligence Trends in Intelligent Systems*, 2015, pp. 156-167.
19. *Jensen J.* A Systematic Literature Review of the Use of Semantic Web Technologies in Formal Education, *Br. J. Edu. Technol.*, 2019, Vol. 50, pp. 505-517.
20. *Gulyakina N.A., Davydenko I.T.* Semanticheskie modeli i metod soglasovannoy razrabotki baz znaniy [Semantic models and the method of coordinated development of knowledge bases], *Programmnye produkty i sistemy* [Software products and systems], 2020, No. 3, pp. 420-429.
21. *Shalimov P.Yu.* Modeli kolichestvennoy shkaly otsenki semanticheskoy informatsii [Models of a quantitative scale for evaluating semantic information], *Avtomatizatsiya i modelirovanie v proektirovanii i upravlenii* [Automation and modeling in design and management], 2020, No. 2 (8), pp. 24-32.

Статью рекомендовал к опубликованию к.т.н., доцент С.Г. Буланов.

Слепцов Алексей Геннадьевич – Южный федеральный университет; e-mail: alslepcov@sfedu.ru; г. Таганрог, Россия; тел.: 88634371651; кафедра систем автоматизированного проектирования; аспирант.

Берешполов Игорь Сергеевич – e-mail: bereshpolov@sfedu.ru; кафедра систем автоматизированного проектирования; аспирант.

Кравченко Юрий Алексеевич – e-mail: yakravchenko@sfedu.ru; кафедра систем автоматизированного проектирования; д.т.н.; доцент.

Sleptsov Aleksey Gennadievich – Southern Federal University; e-mail: alslepcov@sfedu.ru; Taganrog, Russia; phone: +78634371651; the department of computer aided design; postgraduate.

Bereshpolov Igor Sergeevich – e-mail: bereshpolov@sfedu.ru; the department of computer aided design, postgraduate.

Kravchenko Yuriy Alekseevich – e-mail: yakravchenko@sfedu.ru; the department of computer aided design; dr. of eng. sc.; associate professor.

УДК 681.2.089

DOI 10.18522/2311-3103-2023-3-201-211

С.И. Клевцов

ПОРОГОВАЯ ОЦЕНКА СОСТОЯНИЯ ТЕХНИЧЕСКОГО ОБЪЕКТА НА ОСНОВЕ СЕГМЕНТАЦИИ И ИДЕНТИФИКАЦИИ МОДЕЛИ КОНТРОЛИРУЕМОГО ПАРАМЕТРА

Для фиксации скачков среднего значения предложен способ детектирования, основанный на сегментации исследуемого сигнала на основе формирования кумулятивных сумм с использованием критерия Пейдж-Хинкли. Использование критерия правдоподобия Пейдж-Хинкли позволяет в реальном масштабе времени обнаружить скачкообразные изменения среднего значения контролируемого параметра объекта в условиях шума. При использовании метода предполагается, что сигнал описывается временным рядом значений исследуемого сигнала. Из этого ряда можно выделить отдельные последовательные участки, которые можно рассматривать как некоторые модели сигнала, ограниченные по времени. В основе метода лежит использование статистики критерия, на основе которой сравниваются две или три модели, оцененные по различным участкам сигнала, что позволяет обнаруживать скачкообразные изменения в параметрах модели. В методе предполагается, что рассматривается кусочно-постоянный сигнал с аддитивной помехой. В произвольные моменты времени происходят скачки среднего значения данного сигнала. Скачки среднего значения сигнала могут быть разными по знаку (фиксируются по разные стороны от оси времени) и значительно превышают исходное значение по модулю. Среднее значение сигнала – постоянная величина, близкая к нулю. Но возможна ситуация, когда повторный скачок будет сделан с уровня, отличного от близкого к нулю среднего значения, причем как в направлении возрастания, так и убывания среднего значения сигнала и изменения полярности сигнала (знака значений сигнала). Выбран критерий, который позволяет минимизировать время запаздывания в обнаружении скачка среднего значения снимаемого сигнала с минимумом ложных тревог. При этом используется сегментации исследуемого сигнала на основе формирования кумулятивных сумм с использованием критерия Пейдж-Хинкли. Использование критерия правдоподобия Пейдж-Хинкли позволяет в реальном масштабе времени обнаружить скачкообразные изменения среднего значения контролируемого параметра объекта в условиях шума.

Модель; критерий правдоподобия; скачок сигнала; детектирование; ускорение; объект.