УДК 004.89

МЕТОД АВТОМАТИЧЕСКОГО ПРОДЛЕНИЯ ВИДЕОРЯДА НА ОСНОВЕ ВИЗУАЛЬНЫХ И СЕМАНТИЧЕСКИХ СВОЙСТВ ВИДЕОСЦЕН

Вишневская Татьяна Ивановна[1],
кандидат физико-математических наук,
e-mail: iu7vt@bmstu.ru,

Уточкина Наталия Витальевна[1],
e-mail: unatart@outlook.com,
[1]*Московский государственный технический университет (МГТУ им. Н.Э. Баумана), г. Москва, Россия*

В статье представлен метод автоматического продления видеоряда с использованием иерархических рекуррентных нейронных сетей. В основе метода лежит информация о визуальных и семантических свойствах видеосцен, полученных из видео, и заранее известные данные о видео. В статье представлены алгоритмы выделения визуальных и семантических свойств видеосцен, алгоритм продления видеоряда. Дано описание результата работы метода. Продемонстрирована целесообразность использования метода. Описаны возможные сценарии использования метода. Предложенный метод продления видеоряда может быть использован при рекомендации коротких видео-роликов, а также фильмов. В результате работы метода получается следующий набор видео, каждый элемент которого близок по выделенным свойствам к начальной выборке.

Ключевые слова: рекомендации, видео, свойства видеосцен, иерархические нейронные сети, продление видеоряда

A VIDEO-SEQUENCE AUTOMATICAL CONTINUATION METHOD BASED ON THE VIDEO-SCENES VISUAL AND SEMANTIC PROPERTIES

Vishnevskaya T.I.[1],
PhD in Mathematics,
e-mail: iu7vt@bmstu.ru,

Utochkina N.V.[1],
e-mail: unatart@outlook.com,
[1]*Bauman Moscow State Technical University, Moscow, Russia*

This article suggests a continuation method for a video-sequence by means of hierarchical recurrent neural networks. The method is based on the information about the visual and semantic properties of video-scenes obtained from the video and its previously known data. The article presents algorithms for identifying the visual and semantic properties of video- scenes and an algorithm for extending the video sequence as well. It describes the result of the method and demonstrates its worth. It also describes possible scenarios its usage. The method can be used for recommending short video clips and films. As a result a set of videos is obtained where each element is close to the initial selection in terms of the selected properties.

Keywords: recommendations, video, video scene properties, hierarchical RNN, video-sequence continuation

DOI 10.21777/2500-2112-2021-3-53-59

Введение

Рекомендации стали важной функциональной частью практически любого сервиса в интернете. Продление видеоряда – частный случай задачи построения пользовательских рекомендаций для медиа-сервисов.

Видео-контент является неотъемлемой составляющей большинства современных медиа-сервисов. Миллионы пользователей по всему миру ежедневно просматривают кино и сериалы в онлайн-кинотеатрах, короткие ролики на стриминговых платформах и в социальных сетях.

Когда пользователь входит в приложение онлайн-кинотеатра, он видит список контента, который, по мнению рекомендательной системы, понравился бы ему, исходя из истории просмотров и оценок. В рамках одного посещения медиа-сервиса пользователь может посмотреть последовательно несколько видео (особенно если речь идет о коротком контенте). Длительный и непрерывный просмотр видео является косвенным сигналом удовлетворенности пользователя от сервиса [1].

Чтобы обеспечить непрерывный просмотр, пользователю должна быть доступна последовательность видео, которая отвечает сразу нескольким свойствам: когерентности, семантической и стилистической схожести, разнообразию, востребованности, релевантности пользователю и т.д. [2]. В случае, когда на сервисе доступно множество тысяч видеофайлов, решить такую задачу вручную становится невозможным. Для автоматического создания подобных последовательностей используются рекомендательные системы, а сама задача называется «продление плейлиста» (в англ. яз. – playlist continuation).

Целью данной работы является создание метода автоматического продления видеоряда на основе визуальных и семантических свойств видеосцен. Этот метод позволит качественно предлагать следующие видео пользователю, исходя из его истории просмотров.

Научная новизна работы заключается в том, что в работе используется большое количество выделенных свойств видеостен, а рекомендательная система строится на иерархической рекуррентной нейронной сети, в то время как известные методы [3; 4; 5] используют метод коллаборативной фильтрации для рекомендаций или метод опорных векторов. Рекуррентные нейронные сети уже были использованы для рекомендаций и они показывают на 15–30% более качественный результат, чем упомянутые ранее методы [6].

Свойства видеофайлов

В работе используется набор данных из более 1200 трейлеров, а также описание фильмов, к которым трейлеры принадлежат. Описание фильмов состоит из следующих параметров: название фильма, название трейлера, длина фильма, рейтинг фильма, жанры, место и время событий фильма, синопсис. Трейлер, как любой видеофайл, можно разделить на сцены. Каждую сцену можно описать некоторым набором свойств. В таблице 1 описаны полученные свойства из видеосцен и данных файлов, описывающих фильм (в таблице – мета свойства).

Таблица 1 – Свойства видеофайла

Название свойств	Список свойств
Зрительные	Яркость, насыщенность, цвета, темп смены цветов от сцене к сцене, главный цвет, объекты, присутствующие на сцене (например, «человек», «машина»)
Аудио	Энергия, темп, амплитуда, гармоника, MFCC, RMS, полоса пропускания, хроматограмма, «завал» энергии
Мета	Рейтинг фильма, жанр, место и время событий фильма, эмоциональный окрас синопсиса, ключевые слова (например, «преследование», «путешествие», «драконы»)

Алгоритм получения свойств видеосцен

Для получения свойств видеосцен необходимо разделить видеофайл на сцены. Существует множество известных методов для разделения видеофайла на сцены. В работе используется один из базовых методов – при помощи анализа изменения цветового пространства HSV [7]. Далее для каждой из полученных сцен выделяются визуальные характеристики. На протяжении всей работы используется цветовое пространство HSV. Темп смены цветов от сцены к сцене классифицируется от низкого к высокому и считается изменением каждого из параметров цветового пространства HSV – тон, насыщенность, яркость. Стоит учесть, что сцена состоит из нескольких кадров, значения по сцене считаются средним по кадрам, входящим в сцену. Объекты на сцене находятся при помощи модели распознавания объектов ResNet-50. Характеристики аудио так же делятся «посценно» и берутся их средние значения.

Мета информация, за исключением синопсиса берется «как есть». Из синопсиса выделяется эмоциональный окрас описания при помощи нейросети построенной на данных RuSentiment [8]. Эмоциональный окрас содержит оценку по параметрам: позитивность, негативность, нейтральность.

На рисунке 1 отображена схема представления трейлера как совокупности посценных зрительных и аудио данных, эмоционально разобранного синопсиса и метаинформации.

Рисунок 1 – Представление трейлера

На рисунке 2 представлен обобщенный алгоритм выделения зрительных и аудио характеристик из видеофайла и записи их в CSV-файл, который будет использоваться в алгоритме продления видеоряда.

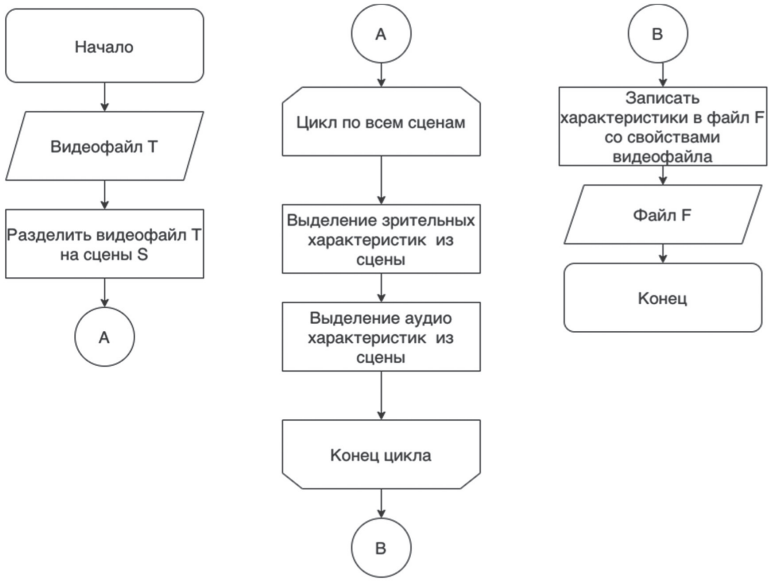

Рисунок 2 – Схема алгоритма выделения свойств из трейлера

Алгоритм продления видеоряда на основе выделенных свойств

Решаемая задача строится следующим образом:

Есть массив сессий пользователей **S** размера *M*, каждый элемент – id просмотренного трейлера: $\mathbf{S} = \{s_1, s_2, s_3, \ldots, s_M\}$, каждый элемент s_i можно описать следующим образом:

1) набором сцен **V** размера **N**, из которого состоит трейлер

$$\mathbf{V} = \{v_1, v_2, \ldots, v_n\},$$

где v_i можно представить в виде набора свойств, описывающих сцену **Q**, размера *L*: $\mathbf{Q} = \{q_1, q_2, \ldots, q_L\}$;

2) набором дополнительных данных размера *K*: $\mathbf{D} = \{d_1, d_2, \ldots, d_K\}$.

Цель задачи состоит в том, чтобы найти следующий элемент сессии пользователя, который можно предложить пользователю.

Данный метод основан на иерархической рекуррентной нейронной сети. На вход подаются следующие компоненты: выделенные ранее посценные характеристики для каждого трейлера, последовательность просмотренных пользователем ранее трейлеров в виде списка идентификаторов (id) трейлеров, а также дополнительная информация о фильме для каждого трейлера.

На выходе ожидается последовательность в виде списка идентификаторов трейлеров, которые предложены пользователю на просмотр. Новая последовательность должна быть когерентна, не должна иметь дубли с входной последовательностью и должна быть семантически и стилистически схожа с входной последовательностью.

Сама модель состоит из двух уровней RNN, в основе которых лежит ячейка GRU – управляемый рекуррентный блок [9]: первый уровень – по сценам каждого из трейлеров, второй уровень – по последовательности трейлеров.

На рисунке 3 и при помощи формул (1)–(5) представлена разработанная модель с описанием входных данных в каждый из компонентов.

Перед тем как подать данные в модель для обучения, необходимо нормализовать данные, преобразовать все столбцы данных, имеющие словарное описание в цифровые значения, очистить данные от нерелевантных значений. Состояние *GRU* по сцене

$$\mathbf{h}_{si} = GRU\left(s_{ij}\right), \tag{1}$$

где s – свойства j-ой сцены i-го трейлера, $j \in [0, M]$, $i \in [0, N]$.

$$\mathbf{f}_i = \mathbf{h}_{sM} \text{ – финальное состояние GRU по сценам.} \tag{2}$$

Состояние GNU по трейлеру

$$\mathbf{h}_i = GRU\left(\mathbf{X} \cdot d_i \cdot f_i\right), \tag{3}$$

где d_i – дополнительные свойства *i*-го трейлера;

X – векторное представление id элемента x_i i-го трейлера.

$$\mathbf{y} = \text{softmax}\left(\mathbf{h}_i\right). \tag{4}$$

$$x_{i+1} = \max\left(\mathbf{y}\right). \tag{5}$$

Полученный рейтинг необходимо подать в фильтр уже просмотренных видео для того, чтобы пользователю они не предлагались. Таким образом, выходное значение рейтинга для всех рассмотренных алгоритмов можно пересчитать с помощью формулы (6).

$$\mathbf{y} = \mathbf{F} \times \mathbf{y}, \text{ где } \mathbf{F} = \{f_1, f_2, \ldots, f_N\},\ f_i \in \{0,1\}, i \in [0, N]. \tag{6}$$

На рисунке 4 представлен алгоритм продления видеоряда на основе семантических и визуальных свойств видеосцен, который включает в себя сбор и подготовку данных, обучение модели и предсказание результатов.

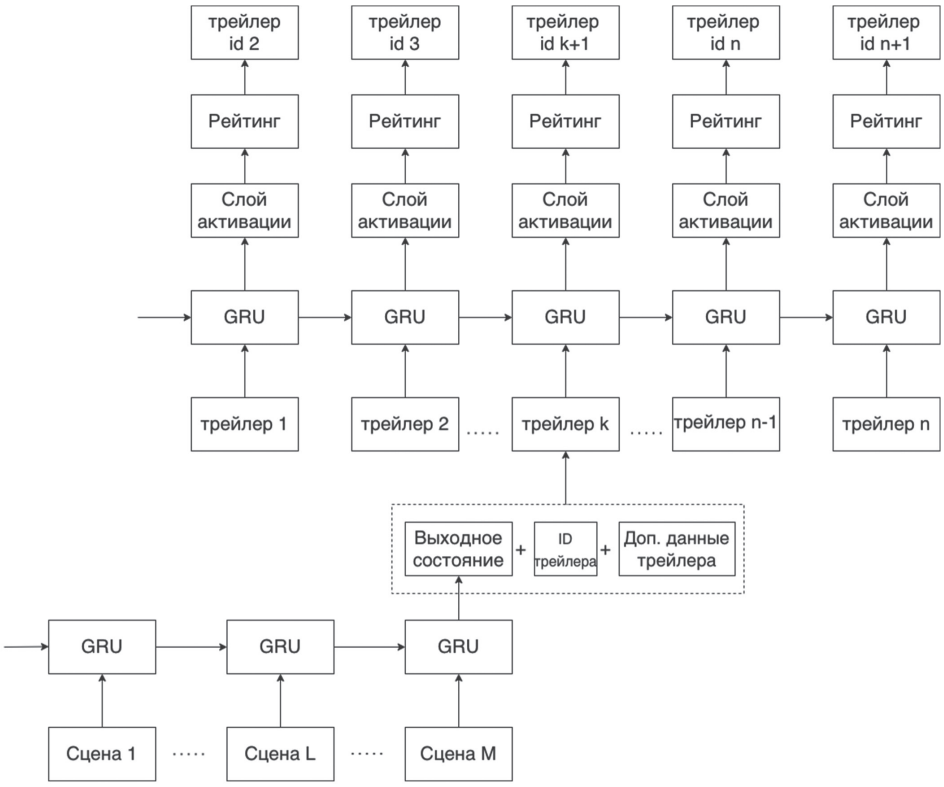

Рисунок 3 – Модель автоматического продления видеоряда

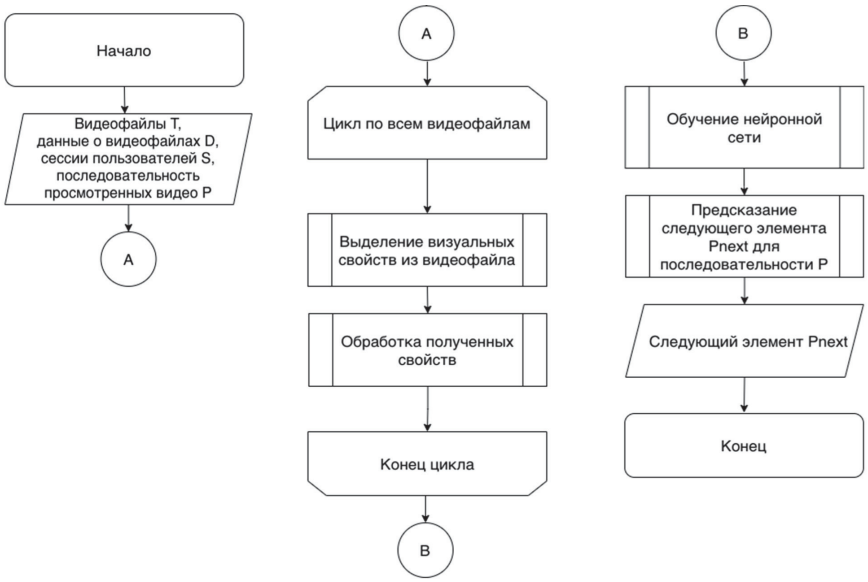

Рисунок 4 – Схема алгоритма автоматического продления видеоряда на основе семантических и визуальных свойств видеосцен

Результат работы разработанного метода

На этапах обучения и тестирования модели получены следующие данные о точности и потерях, представленные на рисунке 5.

В качестве оптимизатора был использован оптимизатор Adam [10], т.к. он был специально спроектирован для использования в моделях глубокого обучения. В качестве функции потерь была исполь-

зована перекрестная энтропия, т.к. выходной слой показывает рейтинг всех имеющихся трейлеров на роль следующего, она описана формулой (7).

$$H(p,q) = -\sum_{x \in \chi} p(x) \log q(x). \quad (7)$$

Видно, что точность и потери прямо пропорциональны друг к другу и полный цикл обучения составляет 100 эпох, после чего функция потерь начинает расти.

Разработанная модель соответствует следующим свойствам:

1) когерентности – просмотренные трейлеры и предложенный трейлер совпадают по визуальным характеристикам;

2) новизне – благодаря фильтру, который находится на последнем слое модели, пользователь не будет получать предложенные трейлеры, которые он недавно посмотрел.

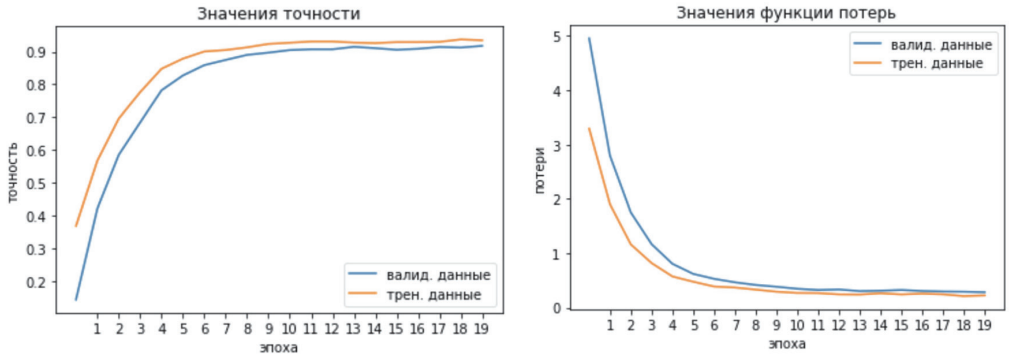

Рисунок 5 – Результаты замеров точности и функции потерь

Заключение

Большинство известных работ описывают процесс построения рекомендаций на основе визуальных и семантических свойств видеосцен с применением коллаборативной фильтрации или метода опорных векторов.Научная новизна данной работы заключается в том, что в ней предлагается метод, использующий в своей основе иерархические рекуррентные сети. Метод был разработан в результате применения подхода нейролингвистического программирования, что позволило проследить аналогию «текст – последовательность слов» и «видео – последовательность сцен». За счет иерархической структуры сети, появляется возможность учитывать тренд изменения параметров сцен отдельно от основных данных видео. Данный подход позволяет качественно учитывать предпочтения пользователя – выдавать контент, схожий с тем, что видел пользователь ранее, с высокой точностью, что подкрепляется результатами исследований в данной работе.

Предложенный метод может иметь практическое значение для построения пользовательских рекомендаций для медиасервисов.

Полученные результаты могут быть полезны разработчикам в сфере машинного обучения.

Список литературы

1. *Catherine, W.* Why TikTok made its user so obsessive? The AI Algorithm that got you hooked / Catherine W., 2020 [Электронный ресурс]. – URL: https://towardsdatascience.com/why-tiktok-made-its-user-so-obsessive-the-ai-algorithm-that-got-you-hooked-7895bb1ab423.

2. *Dietmar, J.* Music Recommendations: Algorithms, Practical Challenges and Applications / Dietmar J., Iman K., Geoffray B. // Collaborative Recommendations, 2018. Pp. 481–518.

3. *Luca, C.* Affective Recommendation of Movies Based on Selected Connotative Features / Luca C., Sergio B., Riccardo L. // IEEE Transactions on Circuits and Systems for Video Technology. – Vol. 23. – No. 4. – Pp. 636–647. – April 2013. DOI: 10.1109/TCSVT.2012.2211935.

4. *Xingzhong, D.* Personalized Video Recommendation Using Rich Contents from Videos / Xingzhong D. [*et al.*] // du2018personalized, 2018.
5. *Yashar, D.* Using visual features based on MPEG-7 and deep learning for movie recommendation. / Yashar D., Mehdi E., Massimo Q., Paolo C. // International Journal of Multimedia Information Retrieval. – 2018. – № 7(4). – Pp. 1–13.
6. *Massimo, Q.* Personalizing Session-based Recommendations with Hierarchical Recurrent Neural Networks / Massimo Q., Balázs H., Alexandros K., Paolo C. // RecSys '17: Proceedings of the Eleventh ACM Conference on Recommender Systems. – 2017. – Pp. 130–137.
7. *Zeeshan, R.* Detection and representation of scenes in videos / Zeeshan R., Mubarak S. // IEEE transactions on Multimedia. – 2005. – № 7(6). – Pp. 1097–1105.
8. *Rogers, A.* RuSentiment: An Enriched Sentiment Analysis Dataset for Social Media in Russian / Rogers A. [*et al.*] // Conference: Proceedings of the 27th International Conference on Computational Linguistics (COLING 2018), 2018, Santa Fe, NM.
9. *Balázs, H.* Session-based recommendations with recurrent neural networks / Balázs H. [*et al.*] // ICLR, 2016.
10. *Kingma D.P., Ba J. Adam.* A Method for Stochastic Optimization // 3rd International Conference for Learning Representations, San Diego, 2015.

Reference

1. *Catherine, W.* Why TikTok made its user so obsessive? The AI Algorithm that got you hooked / Catherine W., 2020 [Электронный ресурс]. – URL: https://towardsdatascience.com/why-tiktok-made-its-user-so-obsessive-the-ai-algorithm-that-got-you-hooked-7895bb1ab423.
2. *Dietmar, J.* Music Recommendations: Algorithms, Practical Challenges and Applications / Dietmar J., Iman K., Geoffray B. // Collaborative Recommendations, 2018. Pp. 481–518.
3. *Luca, C.* Affective Recommendation of Movies Based on Selected Connotative Features / Luca C., Sergio B., Riccardo L. // IEEE Transactions on Circuits and Systems for Video Technology. – Vol. 23. – No. 4. – Pp. 636–647. – April 2013. DOI: 10.1109/TCSVT.2012.2211935.
4. *Xingzhong, D.* Personalized Video Recommendation Using Rich Contents from Videos / Xingzhong D. [et al.] // du2018personalized, 2018.
5. *Yashar, D.* Using visual features based on MPEG-7 and deep learning for movie recommendation. / Yashar D., Mehdi E., Massimo Q., Paolo C. // International Journal of Multimedia Information Retrieval. – 2018. – № 7(4). – Pp. 1–13.
6. *Massimo, Q.* Personalizing Session-based Recommendations with Hierarchical Recurrent Neural Networks / Massimo Q., Balázs H., Alexandros K., Paolo C. // RecSys '17: Proceedings of the Eleventh ACM Conference on Recommender Systems. – 2017. – Pp. 130–137.
7. *Zeeshan, R.* Detection and representation of scenes in videos / Zeeshan R., Mubarak S. // IEEE transactions on Multimedia. – 2005. – № 7(6). – Pp. 1097–1105.
8. *Rogers, A.* RuSentiment: An Enriched Sentiment Analysis Dataset for Social Media in Russian / Rogers A. [et al.] // Conference: Proceedings of the 27th International Conference on Computational Linguistics (COLING 2018), 2018, Santa Fe, NM.
9. *Balázs, H.* Session-based recommendations with recurrent neural networks / Balázs H. [et al.] // ICLR, 2016.
10. *Kingma D.P., Ba J. Adam*. A Method for Stochastic Optimization // 3rd International Conference for Learning Representations, San Diego, 2015.