

УДК 33

DOI: 10.34670/AR.2022.32.31.008

Внедрение новых технологий и IT-решений для телевизионной индустрии

Кузнецов Святослав Сергеевич

Технический директор телеканалов San Porto и Torre Ricca,
ООО «Торре Рикка»,
115533, Российская Федерация, Москва, пр. Андропова, 22;
e-mail: svyatique@gmail.com

Аннотация

В современных условиях общественного развития цифровизация общественных процессов диктует условия внедрения в различные сферы экономики инновационных технологий, в основе которых лежит необходимость решения конкретных задач. Телевизионная индустрия не является тому исключением – появление новых технологий и IT-решений в телевизионной индустрии всегда сопряжено с необходимостью решения конкретных задач, одной из которых является задача сбора и интеллектуальной обработки данных, позволяющей создать эффективную интеллектуальную аналитическую платформу для данных, аккумулируемых в телевизионной индустрии. В статье автором предложено формализованное описание математической модели интеллектуального анализа данных для целей внедрения в базовые процессы обработки данных в телевизионной индустрии, которое дает представление о модели и позволяет в дальнейших исследованиях перейти к ее программной реализации. Используемые в рамках формализованной модели алгоритмы показывают возможности обработки и анализа больших объемов данных, аккумулируемых в сфере телевизионной индустрии. Предложенная формализованная модель интеллектуального анализа данных является основой для реализации эффективного прогнозирования, поскольку полученные в результате анализа данные могут братья в качестве основы для принятия управленческих решений.

Для цитирования в научных исследованиях

Кузнецов С.С. Внедрение новых технологий и IT-решений для телевизионной индустрии // Экономика: вчера, сегодня, завтра. 2022. Том 12. № 6А. С. 69-76. DOI: 10.34670/AR.2022.32.31.008

Ключевые слова

Кластеризация, коэффициент корреляции, обработка данных, новые технологии, интеллектуальный анализ данных.

Введение

В современных условиях экономического развития, сопровождающихся глобальной цифровизацией большинства процессов, использование новых технологий и IT-решений становится необходимым условием цифрового развития. Телевизионная индустрия не является тому исключением – цифровые технологии проникают во все процессы работы телевизионной индустрии, меняя рынок медиапотребления. Так, например, быстрое развитие интернет-индустрии и различных медиатехнологий сместило некогда монопольное положение традиционных средств массовой информации (далее – СМИ), в пользу поддающихся количественной оценке интерактивных электронных каналов. В специализированной литературе отмечается, что новые цифровые технологии значительно сокращают затраты на производство и доставку телевизионного контента, увеличивают скорость доступа к нему, а также существенно повышают степень персонализации телевизионного контента [Барсегян и др., 2009; Cerqueira et al., 2009; Distributed Ledger Technology..., www; Grym, www]. В указанных условиях становится вполне понятной растущая озабоченность компаний, предоставляющих услуги в сфере телевизионной индустрии, необходимостью создания более прочных связей с потребителями услуг.

В последние несколько лет интерактивные электронные каналы, которые именуют также термином «новые медиа», постепенно преодолевают ограничения рынка телевизионной индустрии, заставляя обращать внимание на то влияние, которое оказывают новые технологии и IT-решения на развитие телевизионной индустрии. В указанных условиях проблематика внедрения и совершенствования управления новыми технологиями и IT-решениями в телевизионной индустрии видится весьма актуальной.

Вместе с тем, появление новых технологий и IT-решений в телевизионной индустрии, помимо необходимости цифровизации процессов, всегда сопряжено с необходимостью решения конкретных задач. В ограниченных рамках настоящей статьи акцентируем внимание на методологии внедрения технологий, направленных на решение задач в области сбора и интеллектуальной обработки данных, позволяющих создать эффективную интеллектуальную аналитическую платформу для больших данных, аккумулируемых в телевизионной индустрии.

Основная часть

В научной литературе относительно большого объема данных для кино и телевидения отмечается, что такие данные фокусируются на сети как на информационной платформе, которая относится к информации о массовых данных, генерируемой при создании, передаче и приеме информации, а также системы для хранения, обработки и представления такой информации [Дюк, 2011; Cerqueira et al., 2009].

Такие большие массивы информации должны подвергаться качественному анализу. На основе фактов и предположений, почерпнутых из открытых источников, можно не только анализировать состояние телевизионной индустрии в конкретный момент времени, но и строить прогнозы развития ситуации, что важно для принятия эффективных управленческих решений. В этой связи технологические решения, позволяющие осуществлять интеллектуальный анализ данных, являются одним из наиболее перспективных направлений, нацеленных на решение практических проблем телевизионной индустрии, когда речь заходит о внедрении технических решений, основанных на искусственном интеллекте и аккумулировании информации.

Специалисты отмечают, что преимущества интеллектуального анализа данных позволяют выявлять скрытую, ранее неизвестную и потенциально ценную информацию из больших объемов данных [Бахитова, Исламов, 2012; 1, Фомина, 2021; Distributed Ledger Technology..., www]. Интеллектуальный анализ данных также является своего рода процессом поддержки принятия решений [Комаров, 2019; Psannis, 2018]. Основанный главным образом на искусственном интеллекте, машинном обучении, а также на распознавании образов, статистике, базах данных и технологиях визуализации, интеллектуальный анализ данных позволяет в автоматическом режиме извлекать всю необходимую информацию для принятия эффективных управленческих решений в телевизионной индустрии.

Методологически процесс принятия решений будет разделен на несколько важнейших этапов: подготовка данных, необходимых для проведения анализа; поиск правил осуществления анализа или же сам процесс интеллектуального анализа данных; проверка и интерпретация результатов; принятие итогового решения.

Этап подготовки данных включает в себя отбор необходимых данных из соответствующего источника данных и их интеграцию в набор данных, который будет использоваться для интеллектуального анализа данных. На втором этапе – этапе осуществления анализа – обрабатываются отобранные данные с помощью определенных методов или алгоритмов. Третий этап – проверка результатов – требует изучения обработанных данных, а интерпретация результатов заключается в предоставлении интуитивных, разумных и понятных описаний с использованием логических рассуждений.

Поскольку целью интеллектуального анализа данных является извлечение наиболее полезных данных, *процесс требует сочетания извлечения, анализа, преобразования, классификации, организации, обоснования выбора и других процессов, связанных с обработкой данных. Вышеназванные процессы представляются сложной задачей, особенно для сферы телевизионной индустрии, поскольку в ней существует слишком много типов данных и они объемны, кроме того, они достаточно быстро устаревают. Вместе с тем, традиционные методы обработки и возможности анализа данных в телевизионной индустрии больше не могут удовлетворять спрос.* Поэтому, как уже было отмечено ранее, наиболее оптимальным технологическим решением является решение, которое позволит создать эффективную интеллектуальную аналитическую платформу для данных, обрабатываемых в телевизионной индустрии.

Наиболее оптимальным при этом видится решение, в основе которого лежит кластерный анализ данных. Кластерный анализ представляет собой процесс классификации похожих или идентичных объектов по нескольким категориям. Иными словами, для анализа собираются похожие наборы данных для классификации. Поэтому k-средства в алгоритме кластерного анализа используются для анализа сходства того или иного объекта, например, телевизионных передач с высоким рейтингом среди телевизионных передач различной направленности. Серия выборок объединяется в k различных категорий (где k – количество категорий модели), то есть фактическая цель модели. Функция будет представлять собой сумму различий между классами. Целью кластеризации K-средних является минимизация дисперсии и начальных значений всех классов. K центры классов – представляют собой центральные точки класса. При повторной итерации получается модель, которая будет оптимально отвечать всем требованиям, а конечным условием итерации является максимальное количество итераций и значение сходимости [Cerqueira et al., 2009].

Теперь акцентируем внимание на специфике предварительной обработки данных, которую можно использовать в рамках предложенной методологии работы с анализом данных. Обычно объем данных в сети огромен, кроме того, наблюдается слишком много источников получения данных, что может привести к получению ненадежных данных для осуществления интеллектуального анализа. Так, например, чтобы собрать информацию о телевизионной передаче, необходимо получить видеоданные и пользовательские данные, которые включают в себя название телевизионной передачи, дату ее выпуска, тип, область производства, режиссера, участников, информацию о популярности и рейтингах. Данные плохого качества приведут к ненадежности анализа данных и большому разрыву с фактическими данными. Осложняет проблему и тот факт, что данные, аккумулируемые из различных источников, отличаются форматами. Поэтому с точки зрения методологии процессу обработки данных должен предшествовать процесс очистки данных, в рамках которого осуществляется: удаление дублирования данных, интеграцию данных и стандартизацию данных. Процесс подготовки данных занимает больше половины времени, которое затрачивается на осуществление интеллектуального анализа данных.

Теперь составим конкретную математическую модель процесса обработки данных, в рамках которой для устранения некорректных данных используем среднее значение и дисперсию.

$$\begin{aligned}\frac{\partial f_i(r)}{\partial r_i} &= \frac{\exp(r) \sum_i \exp(r_j) - \exp^2(r_i)}{\sum_j \exp(r_i)} = \\ &= \frac{\exp(r_i) \sum_i (r_i)}{\sum_j \exp(r_j)} * \frac{\exp(r_i) \sum_i \exp(r_j) - \exp(r_i)}{\sum_j \exp(r_j)} = \\ &= \frac{\exp(r_i)}{\sum_j (r_j)} * \left(1 - \frac{\exp(r_i)}{\sum_j \exp(r_j)}\right) = f_i(r)(1 - f_i(r))\end{aligned}$$

В представленном уравнении за r – это значение всех собранных данных, а f – не имеющее объема (безразмерное) значение данных.

Теперь для устранения различий в размерах между характеристиками различных показателей оценки, необходимо выбрать уравнение для стандартизации данных, при этом для ограничения диапазона различных характеристик до $[0, 1]$, будем использовать максимальные и минимальные методы стандартизации.

$$N_{x,y} = \sqrt{\frac{1}{mn}} \sum_{i=1}^m \sum_{j=1}^n [E_{x,y}(i,j) - \overline{E_{x,y}}]$$

В представленном уравнении N – это значение собранных данных, а E – это данные, приведенные к стандартному формату (нормализованные данные).

Теперь непосредственно акцентируем внимание на обработке данных, для чего выберем уравнение, наиболее подходящее для стратегии обработки данных. Ввиду специфики данных, которые подлежат обработке, наиболее оптимальным математическим инструментарием видится ассоциативный анализ. В специализированной литературе отмечается, что

ассоциативный анализ является идеальным инструментом для поиска причинно-следственных связей между объектами и наборами элементов при обработке большого объема данных [Шилова, 2015]. Расчет коэффициента корреляции R осуществляется исходя из следующей корреляционной функции:

$$R = \frac{E_{x,y}}{\sqrt{e_{xy}e_{yx}}} = \frac{z_{xy}}{z_{xx}z_{yy}} = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}},$$

$$z_{xx} = \sum_{i=1}^n (x - \bar{x})^2 = \sum_{i=1}^n x^2 - \frac{(\sum_{i=1}^n x)^2}{n},$$

$$z_{yy} = \sum_{i=1}^n (y - \bar{y})^2 = \sum_{i=1}^n y^2 - \frac{(\sum_{i=1}^n y)^2}{n},$$

$$z_{xy} = \sum_{i=1}^n (x - \bar{x})(y - \bar{y}) = \sum_{i=1}^n xy - \frac{(\sum_{i=1}^n x)(\sum_{i=1}^n y)}{n}$$

где z представляет собой коэффициент чувствительности, а значение этого коэффициента находится между интервалом $[0, 1]$. Коэффициент корреляции Пирсона (R) характеризует наличие связи между признаками, где x – значение в выборке x , а y – значение в выборке y . При этом $\bar{x}$ – это средняя по x , а $\bar{y}$ – средняя по y .

Степень, сила или теснота корреляционной связи определяются по величине коэффициента корреляции. Сила связи не зависит от её направленности и определяется по абсолютному значению коэффициента корреляции. Максимально возможное абсолютное значение коэффициента корреляции $R_{xy} = 1$; минимальное, $R_{xy} = 0$ [там же].

Оценка статистической значимости коэффициента корреляции R_{xy} осуществляется при помощи t -критерия, рассчитываемого по следующей формуле:

$$t_R = \frac{R_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy}^2}}$$

Полученное значение t_R сравнивается с критическим значением при определенном уровне значимости и числе степеней свободы $n-2$. Если t_R превышает $t_{\text{критическое}}$, то делается вывод о статистической значимости выявленной корреляционной связи [там же].

Оценка степени корреляции будет осуществляться исходя из следующих уравнений с уже известными переменными:

$$\gamma_{A(E)(xy)} = \frac{C^{\hat{F}v} E(xy)}{\widehat{\delta_{Ey}^2} \widehat{\delta_{Ex}^2}},$$

$$\gamma_{A(N)(xy)} = \frac{C^{\hat{F}v} N(xy)}{\widehat{\delta_{Ny}^2} \widehat{\delta_{Nx}^2}},$$

$$\gamma_{A(E+N)(xy)} = \frac{C^{Fv} E(xy) + C^{Fv} N(xy)}{\sqrt{(\widehat{\delta_{Ey}^2} + \widehat{\delta_{Ex}^2})(\widehat{\delta_{Sy}^2} + \widehat{\delta_{Sx}^2})}},$$

Использование предложенной математической модели для интеллектуального анализа данных в телевизионной индустрии позволит осуществлять оценку данных на основании ключевых показателей. Так, например, для телевизионных передач в качестве них будут – название телевизионной передачи, тематика (жанр) телевизионной передачи, регион выпуска, объем вложенных средств, рейтинг в сети, целевая аудитория, продолжительность выпуска, популярность у телезрителей и другие ключевые слова. Аналогичным образом можно собирать и анализировать данные по различным фильмам, сериалам. Анализ ассоциаций заключается в поиске закона между соответствующими данными о передаче и ее зрительской оценкой.

Заключение

Подводя итог настоящему исследованию, необходимо отметить, что автором предложено формализованное описание математической модели интеллектуального анализа данных для целей внедрения в базовые процессы обработки данных в телевизионной индустрии, которое дает представление о модели и позволяет в дальнейших исследованиях перейти к ее программной реализации. Используемые в рамках формализованной модели алгоритмы показывают возможности обработки и анализа больших объемов данных, аккумулируемых в сфере телевизионной индустрии. Предложенная формализованная модель интеллектуального анализа данных является основой для реализации эффективного прогнозирования, поскольку полученные в результате анализа данные могут братья в качестве основы для принятия управленческих решений.

Библиография

1. Барсегян А.А. и др. Анализ данных и процессов. СПб.: БХВ-Петербург, 2009. 512 с.
2. Бахитова Р.Х., Исламов И.Я. Применение математических методов для повышения рейтингов телевидения на медиарынке (на примере телеканала БСТ) // Фундаментальная математика и ее приложения в естествознании. Уфа, 2012. С. 266-267.
3. Дюк В.А. Применение технологий интеллектуального анализа данных в естественнонаучных, технических и гуманитарных областях // Известия российского государственного педагогического университета им. А.И. Герцена. 2011. № 138. С. 77-87.
4. Комаров В.Н. Разработка архитектуры системы мониторинга и анализа публикаций в сети интернет // Передовые инновационные разработки. Перспективы и опыт использования, проблемы внедрения в производство. Казань: Конверт, 2019. Ч. 2. С. 27-29.
5. Фомина А.Н. Совершенствование управления инновациям в сфере телевизионной индустрии в условиях цифровой трансформации // Гуманитарные, социально-экономические и общественные науки. 2021. № 4-2. С. 206-211.
6. Шилова З.В. Теория вероятностей и математическая статистика. Киров, 2015. 158 с.
7. Psannis K.E. et al. Advanced media-based smart big data on intelligent cloud systems // IEEE Transactions on Sustainable Computing. 2018. Vol. 4. No. 1. P. 77-87.
8. Cerqueira E. et al. International Seminar on Future Multimedia Networks. Springer, 2009. P. 242-247.
9. Distributed Ledger Technology: beyond block chain. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/492972/gs-16-1-distributed-ledger-technology.pdf
10. Grym A. Kryptovarät eivät ole rahaa. URL: <https://www.eurojatalous.fi/fi/blogit/2018/kryptovarät-eivat-ole-rahaa/>

Introduction of new technologies and IT solutions for the television industry

Svyatoslav S. Kuznetsov

Technical Director of San Porto and Torre Ricca TV Channels,
Torre Ricca Limited Liability Company,
115533, 22, Andropova ave., Moscow, Russian Federation;
e-mail: svyatique@gmail.com

Abstract

In modern conditions of social development, digitalization of social processes dictates the conditions for the introduction of innovative technologies into various spheres of the economy, which are based on the need to solve specific tasks. The television industry is no exception to this, the emergence of new technologies and IT solutions in the television industry is always associated with the need to solve specific tasks, one of which is the task of collecting and intelligent data processing, allowing to create an effective intelligent analytical platform for data accumulated in the television industry. In the article, the author offers a formalized description of a mathematical model of data mining for the purposes of implementation into the basic data processing processes in the television industry, which gives an idea of the model and allows for further research to proceed to its software implementation. The algorithms used in the framework of the formalized model show the possibilities of processing and analyzing large amounts of data accumulated in the television industry. The proposed formalized model of data mining is the basis for the implementation of effective forecasting, since the data obtained as a result of the analysis can be taken as a basis for making managerial decisions.

For citation

Kuznetsov S.S. (2022) Vnedrenie novykh tekhnologii i IT-reshenii dlya televizionnoi industrii [Introduction of new technologies and IT solutions for the television industry]. *Ekonomika: vchera, segodnya, zavtra* [Economics: Yesterday, Today and Tomorrow], 12 (6A), pp. 69-76. DOI: 10.34670/AR.2022.32.31.008

Keywords

Clustering, correlation coefficient, data processing, new technologies, data mining.

References

1. Bakhitova R.Kh., Islamov I.Ya. (2012) *Primenenie matematicheskikh metodov dlya povysheniya reitingov televideniya na mediarynke (na primere telekanala BST)* [Application of mathematical methods to improve television ratings in the media market (on the example of the BST TV channel)]. In: *Fundamental'naya matematika i ee prilozheniya v estestvoznanii* [Fundamental mathematics and its applications in natural sciences]. Ufa.
2. Barsegyan A.A. et al. (2009) *Analiz dannykh i protsessov* [Analysis of data and processes]. St. Petersburg: BKhV-Peterburg Publ.
3. Cerqueira E. et al. (2009) *International Seminar on Future Multimedia Networks*. Springer.
4. *Distributed Ledger Technology: beyond block chain*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/492972/gs-16-1-distributed-ledger-technology.pdf [Accessed 06/06/2022]
5. Dyuk V.A. (2011) *Primenenie tekhnologii intellektual'nogo analiza dannykh v estestvennonauchnykh, tekhnicheskikh i gumanitarnykh oblastyakh* [Application of data mining technologies in the natural sciences, technical and humanitarian

-
- areas]. *Izvestiya rossiiskogo gosudarstvennogo pedagogicheskogo universiteta im. A.I. Gertsena* [Proceedings of the Russian State Pedagogical University], 138, pp. 77-87.
6. Fomina A.N. (2021) Sovershenstvovanie upravleniya innovatsiyam v sfere televizionnoi industrii v usloviyakh tsifrovoi transformatsii [Improving the management of innovations in the television industry in the context of digital transformation]. *Gumanitarnye, sotsial'no-ekonomicheskie i obshchestvennye nauki* [Humanitarian, socio-economic and social sciences], 4-2, pp. 206-211.
7. Grym A. *Kryptovarat eivät ole rahaa*. Available at: <https://www.eurojatalous.fi/fi/blogit/2018/kryptovarat-eivat-ole-rahaa/> [Accessed 06/06/2022]
8. Komarov V.N. (2019) Razrabotka arkhitektury sistemy monitoringa i analiza publikatsii v seti internet [Development of the architecture of the system for monitoring and analyzing publications on the Internet]. In: *Peredovye innovatsionnye razrabotki. Perspektivy i opyt ispol'zovaniya, problemy vnedreniya v proizvodstvo* [Advanced innovative developments. Prospects and experience of use, problems of implementation in production]. Kazan: Konvert Publ. Part 2.
9. Psannis K.E. et al. (2018) Advanced media-based smart big data on intelligent cloud systems. *IEEE Transactions on Sustainable Computing*, 4, 1, pp. 77-87.
10. Shilova Z.V. (2015) *Teoriya veroyatnostei i matematicheskaya statistika* [Theory of Probability and Mathematical Statistics]. Kirov.
-