Risk, Analysis and Evaluation

Translated Article†

THE METHOD FOR FORECASTING BOX-OFFICE GROSSES OF MOVIES WITH NEURAL NETWORK

Leonid N. YASNITSKII
Perm State National Research University, Perm, Russian Federation
yasn@psu.ru
Corresponding author

Natal'ya O. BELOBORODOVA
Higher School of Economics, Perm, Russian Federation
natasha09.12@mail.ru

Ekaterina Yu. MEDVEDEVA
Higher School of Economics, Perm, Russian Federation
win.mail.ru95@inbox.ru

Abstract
Importance The article focuses on the neural network forecasting in the film-making industry.
Objectives The article examines what opportunities economic and mathematical modeling provides to forecast revenue and profit from coming movie distribution and identifies factors that determine whether film-making business becomes a commercial success.
Methods The economic and mathematical model relies upon the neural network trained with available historical data on movie distribution and including 20 input parameters. Computer experiments were performed with the 'freezing' method. We used the neural network for computations if any of input data changes, meanwhile the rest of them remain the same.
Results Root-mean-square relative error of the model accounted for 13.8 percent, with the coefficient of determination being 0.86 percent. We refer to *The Da Vinci Code*, *Star Wars* to demonstrate what the model is capable of.
Conclusions and Relevance A virtual increase in the film budget influences projections of box-office grosses and revenue differently. Other aspects of films also have an effect on the film-making success. Having conducted computer experiments, we provided our recommendations, which could boost box-office grosses of films. The proposed economic and mathematical model can be used to optimize financial costs and choose parameters to plan new films to come. The model allows for forecasting box-office grosses and profit from film-making, and examines how various aspects influence the commercial result of film-making.

© Publishing house FINANCE and CREDIT, 2017

The editor-in-charge of this article was Irina M. Komarova
Authorized translation by Irina M. Komarova

Introduction

Scholars and managers of large film production companies thoroughly investigate the ways box office grosses can be forecasted using the mathematical modeling method[1] [1, 2]. According to the renowned film analyst, predicting the precise movie revenue before its release is one of the most difficult and important tasks for film-makers [3].

Another expert states that some analysts mainly get intrigued by difficulty and uncertainty in predicting the demand for a motion picture. This unpredictability exposes film-making to risks and makes researchers explore various techniques for forecasting the post-production effect of a motion picture [4].

Such techniques are difficult to set up due to multiple reasons that may influence the film success. Assuming this in their researches, scholars scrutinize various aspects and features of movies.

In 2006, applying the mathematical framework of discriminant analysis, logic regression, classification and regression trees (CART) and neural networks, scholars involved seven independent traits, with the three of them providing the highest effect, i.e. the number of screens, visual effects (VFX) and the cast's fees. The economic and mathematical models resulted in inaccuracies and errors of about 36.9 percent [3].

As researchers from the Christopher Newport University discovered[2], the box office figures were very much impacted by the expected sequel of a film, genre and budget.

The authors [5], who tried to use a small number of easily accessible criteria, found that the animation studio had the biggest impact on the movie revenue.

Some researchers attempted to forecast box-office grosses from perspectives of the behavioral model. The behavioral model implies parameters that influence the behavior of various groups of consumers. Based on the model, the authors of the article, *Predicting Movie Revenue is a Road to Success* referred herein[3], highlight the importance of information about the movie.

In another article [6], researchers suggest using their own parameter – viewers' emotions – so to frame economic and mathematical models.

It is common to consider the amount of box-office revenue as the main criteria for evaluating the film success, since the box-office revenue gives a true view of the financial benefit from the distribution. In the age of information technologies the Internet accumulates various data and traces ongoing changes and updates. It helps evaluate any film and check its rating.

A rating is, definitely, a biased parameter. However, as compared with box-office grosses, this indicator remains as it is, notwithstanding any economic fluctuations, such as inflation, default, GDP growth or fall, etc. Hence, a rating can be used as an interim variable leading us to the correct result.

However, researches [7] based on methods of regression and correlation analysis discovered that the nexus between the film popularity (number of votes in IMDb[4]) and box-office grosses was rather frail. It means that the film popularity depends on various aspects.

For instance, the motion picture may win the audience as a result of an effective advertising campaign or any award. It makes us believe this criterion is unreasonable to include into parameters of the model designated for predicting box-office revenues.

Other researchers [8] made a convincing revelation that the professional level of the cast, age recommendations

[†]For the source article, please refer to: Ясницкий Л.Н., Белобородова Н.О., Медведева Е.Ю. Методика нейросетевого прогнозирования кассовых сборов кинофильмов. Финансовая аналитика: проблемы и решения. 2017. Т. 10. Вып. 4. С. 449–463. URL: https://doi.org/10.24891/fa.10.4.449

[1] Noakk N.V., Nevolin I.V., Tatarnikov A.S.[Method for predicting revenue from rental movies]. *Finansovaya analitika: problemy i resheniya = Financial Analytics: Science and Experience*, 2012, vol. 5, iss. 48, pp. 17–24. (In Russ.); Tatarnikov A.S. [Methods for predicting box-office grosses]. *Byulleten' kinoprokatchika = Booker's Bulletin*, 2012, no. 10-11, pp. 50–56. (In Russ.); Noakk N.V., Znamenskaya A.N. [Analysis of filmgoer emotions as predictive provision of box office]. *Natsional'nye interesy: prioritety i bezopasnost' = National Interests: Priorities and Security*, 2014, vol. 10, iss. 16, pp. 58–66. (In Russ.)

[2] Christopher Newport University. URL: http://cnu.edu

[3] Vedernikov P., Chirikov I. [Predicting movie revenue is a road to success]. *Menedzher kino = Cinema Manager*, 2008, no. 43, pp. 19–20.

[4] Internet Movie Database (IMDb) is the largest international database of information on movies.

and genre proved to be the principal criteria that determined the box-office revenue of a film.

There is an opinion [9] that sequels[5] generate more substantial box-office grosses, since viewers are relatively cognizant of the content and have better predilection for watching the film, rather than going to the cinema for an unknown plot.

Forecasting Techniques

Constructing the economic and mathematical model for predicting box-office grosses, we regard the findings mentioned in the literature underlying the research. Input parameters of the neural network comprise only those criteria that do not take much time for their assessment and can be processed immediately without much labor input.

The training set consists of 168 films with box-office grosses ranging from USD 1 million to 3 billion, with 10 percent of them being included into the test set. We sourced relevant data on movies from www.kinopoisk.ru. Thus, we construct our economic and mathematical model based on neural networks using the following input parameters and system for coding their values:

x_1 is the film release year;

x_2 is the producing country: 1 – USA, 2 – USA in collaboration with other countries;

x_3 is the film director's sex: 1 – male, 2 – female;

x_4 is the basis for the plot: 1 – true story, 2 – literary work (e.g. novel, detective fiction, etc.) or a remake, 3 – the plot draws upon neither true story, nor any literary work, i.e. the idea is the original concept of the scriptwriter, 4 – the film is a parody;

x_5 is the film budget, million USD;

x_6 is age recommendations (in accordance with the Russian content rating system: 1 – 0+, 2 – 6+, 3 – 12+, 4 – 16+, 5 – 18+;

x_7 is the involvement of fictional characters: 0 – no, 1 – talking animals or objects, 2 – robots, 3 – inhabitants of other planets, 4 – magic creatures, 5 – vampires,

werewolves, demons, 6 – those raised from the dead (ghosts, zombies), 7 – superheroes, 8 – several categories of fictional characters;

x_8 is a malicious character: 0 – no, 1 – yes;

x_9 is the film duration, minutes;

x_{10} is the film director's record of successful movies before this one: 0 – no, 1 – yes;

x_{11} is the film director's age as of the film production date;

x_{12} is the film director's awards, like Oscar and/or Golden Globe: 0 – no, 1 – yes;

x_{13} is the Golden Raspberry award held by the film director: 0 – no, 1 – yes;

x_{14} is the film director's nominations for the Oscar and/or Golden Globe awards: 0 – no, 1 – yes;

x_{15} is the cast's nominations for the Oscar and/or Golden Globe awards: 0 – no, 1 – yes;

x_{16} is a type of dramatic genre: 0 – no, 1 – tragedy, 2 – comedy, 3 – drama, 4 – soap opera, 5 – tragicomedy, 6 – genre blend;

x_{17} is an adventure movie, i.e. action films, westerns, thrillers, gangster movies: 0 – no, 1 – yes;

x_{18} is a fiction movie, i.e. fantastic fiction, fantasy films: 0 – no, 1 – yes;

x_{19} is a continuation of another film, i.e. a following series of the franchise: 0 – no, 1 – yes;

x_{20} is a part of a trilogy, where the third part is divided into two films: 0 – no, 1 – yes.

The output parameter is represented with the amount of global box-office grosses (distribution revenue) in million USD.

In addition to film details (x_1, x_2, x_4–x_9, x_{16}–x_{21}), our model regards details of the film director and the cast (x_3, x_{10}–x_{15}). As many film analysts believe, the success of a film and subsequent box-office revenue directly depend on the public opinion on the giftedness of the film director and cast, that is usually perceived through the Oscar and Golden Globe nominations and awards.

---

[5] The continuation of a book, novel or a movie that is produced following the success of a literary work or initial movie.

This idea can be corroborated with the article on *Gone with the Wind*[6]. According to this article, it is Sidney Howard, the playwright and director, who made the film a success. His talent and public affection played one of the major parts in the tremendous release of the film.

The film director's age x_{11} is a new parameter of the proposed model as an indicator of certain experience in life, rather than film making only.

Designing, optimization, training, testing of the neural network and experiments with the mathematical model were guided by the techniques invented by the Perm Scientific School of Artificial Intelligence.

The optimal structure of the neural network constitutes a perceptron (*Fig. 1*) that has twenty input neurons, one hidden layer of six neurons, and one output neuron. We use the hyperbolic tangent function as activation functions of neurons in the hidden and output layers, and the resilient propagation algorithm as the learning algorithm.

In training and testing the neural network, we identified outliers using special techniques [10, 11]. Following the techniques, we excluded examples, one by one, out of the training set and observed the error of the neural network trained with those thinned sets. If an example of the training set constitutes an outlier and deviates from the general pattern immanent in the subject under study, its exclusion out of the training set will reduce the error of the network training and increase its generalizing properties, meanwhile the exclusion of ordinary examples does not have a substantial impact on the network quality.

The quality of the neural network was evaluated using the root-mean-squared relative error, which accounted for 13.8 percent in the training set, with the coefficient of determination being equal to 0.86.

Whereas examples of the training set were not used to train the neural network, we could state it learnt patterns of the modeled subject and could be used for the calculus.

Experimental Part

After we test the neural network with examples and subsequently verify the mathematical model of the neural network to be adequate, we commence our research. The trained neural network responds to changes in input variables and behaves like the subject area would do.

The significance of input parameters is the first aspect that can be explored using the model. The extent to which they influence the modeling outcome represents the amount of box-office grosses.

This influence can be reasonably assessed through the existing technique[7] using the same neural network. We exclude input parameters one by one and observe the testing error. The higher the testing error is, the more significant the excluded parameter is. The resultant histogram is displayed in *Fig. 2*.

The height of the columns means the testing error computed when we exclude the input parameter denoted under the column. That is why it can be construed as the significance of the parameter.

As showcased in the figure, the film budget turns out to be the most important parameter followed by the duration and the parameter called *Is the film produced as a part of the franchise?*

Neural network modeling not only allows to make forecasts, but also conduct visual computer experiments with models [11, 12], adjust forecasts so to support the film making industry. Thus, trying different input parameters of the trained neural network, we set out definite recommendations for raising box-office grosses.

The first series of the experiments with the mathematical model of the neural network sorted out details of *The Da Vinci Code*. This film has a relatively small budget, and its box-office grosses hardly ever set off the costs. In *Fig. 3* (and further), the dark column depicts the factual amount of box-office grosses and lighter ones indicate the amount of virtual box-office grosses, which the film would have received, if its budget is gradually increased by USD 5–30 million.

As seen in *Fig. 4*, box-office grosses of *The Da Vinci Code*, as predicted by the neural network, grow if its budget virtually rises as well. The same is true for contingent profit from the distribution, which is assessed as

---

[6] Davydenko V. [Seven success drivers of the Gone with the Wind]. *Rossiiskaya gazeta*, 2014, February 1.

[7] Yasnitskii L.N. *Vvedenie v iskusstvennyi intellekt* [An introduction to artificial intelligence]. Moscow, Akademiya Publ., 2005, 176 p.

the difference between box-office grosses (revenue) and the film budget.

Fig. 5 unveils the influence of the following significant parameter, i.e. duration. Bigger dots highlight box-office grosses that match the real duration of the film. Projected box-office grosses grow as the duration of *The Da Vinci Code* is extended, though this correlation is nonlinear.

Analyzing *Fig. 6*, we see box-office grosses of *The Da Vinci Code* would be much higher if the film were continued.

We compiled our recommendations, which are believed to boost the box-office grosses of *The Da Vinci Code* (*Table 1*).

According to the data, box-office grosses would demonstrate a 25-percent growth, if the budget were additionally propped up with USD 15 million and the duration was extended for three minutes. The franchise continuation would also add 25 percent.

Another film we examine with our forecasting technique is *Star Wars. Episode 1: The Panthom Menace*. This film attracts our attention since its box-office grosses almost reached USD 1 billion, notwithstanding its modest budget.

According to our estimates (*Fig. 7* and *8*), box-office grosses would rise by USD 5 million, if the film budget were virtually increased by the same amount. However, the profit would remain unchanged. A subsequent increase in the budget would have no effect on the box-office grosses, but driving the profit down. In the mean time, even if the budget saw a minor virtual increase, it would considerably drain the box-office grosses and profit. Thus, we believe film planners chose the optimal amount of its budget.

As reported in *Fig. 9*, neither involvement of other countries, nor a longer duration, nor a virtual increase in the film director's age, nor splitting the film in two parts would boost its box-office grosses. Hence, we get even more reassured that the other parameters of the film prove their best reasonableness, thus making the film a commercial success.

Conclusion

We designed the economic and mathematical model based on the neural network to predict box-office grosses and profit from distribution of motion pictures. The model can be used in the film making industry for decision-making purposes and attaining as high recoupment as possible.

Table 1
Recommendations for *The Da Vinci Code* movie

Parameter	Source Data	Recommendation
The film production year	2006	2006
Producing country	2	2
Film director's sex	1	1
The basis for the plot	2	2
Film budget	125	140
Age recommendations	3	3
Fictional characters	0	0
Malicious character	1	1
Film duration	149	152
Film director's record of successful films	1	1
Film director's age	52	52
Awards held by the film director	1	1
The Golden Raspberry Award held by the film director	0	0
Film director's nominations for awards	1	1
The cast's nominations for awards	1	1
Type of dramatic genre	0	0
Adventure film	1	1
Fantastic fiction film	0	0
Part of a franchise	0	1
The split of the third part of the film	0	0
Box-office grosses of the film	758	960

Source: Authoring
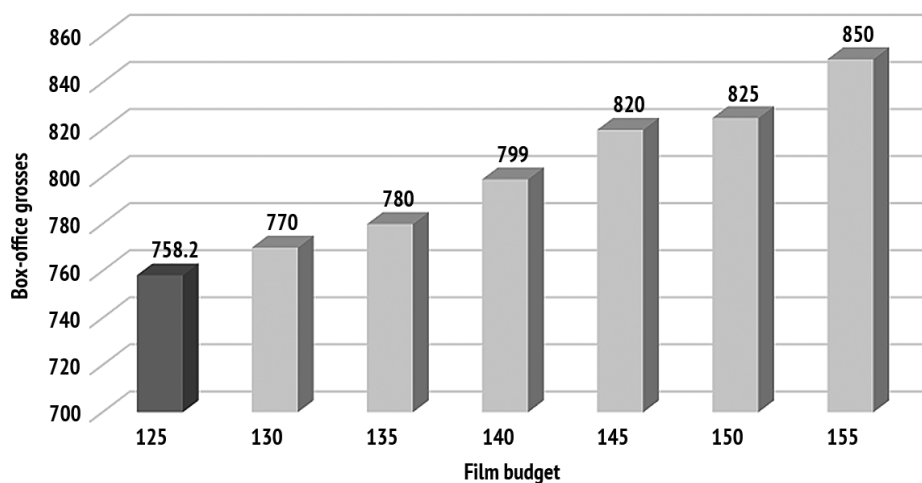
Figure 1
Perceptron
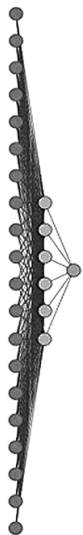
Source: Authoring

Figure 2

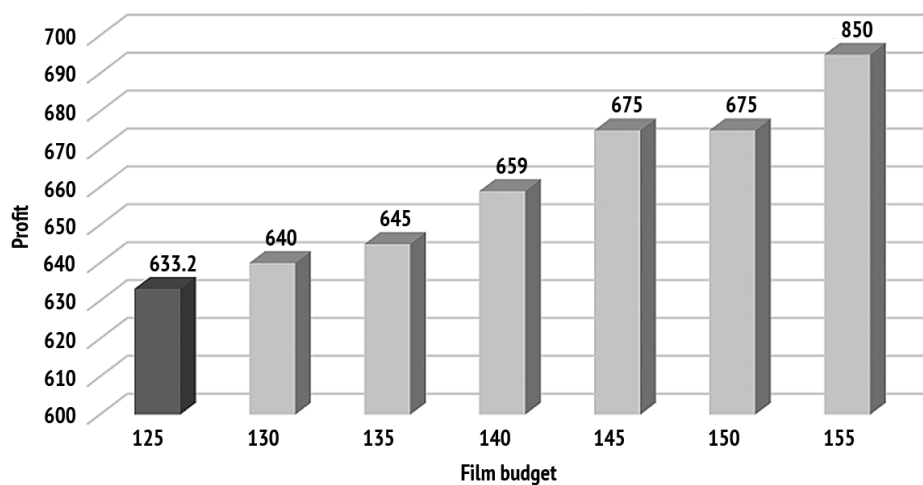**Significance of indicators of box-office grosses for U.S. movies**
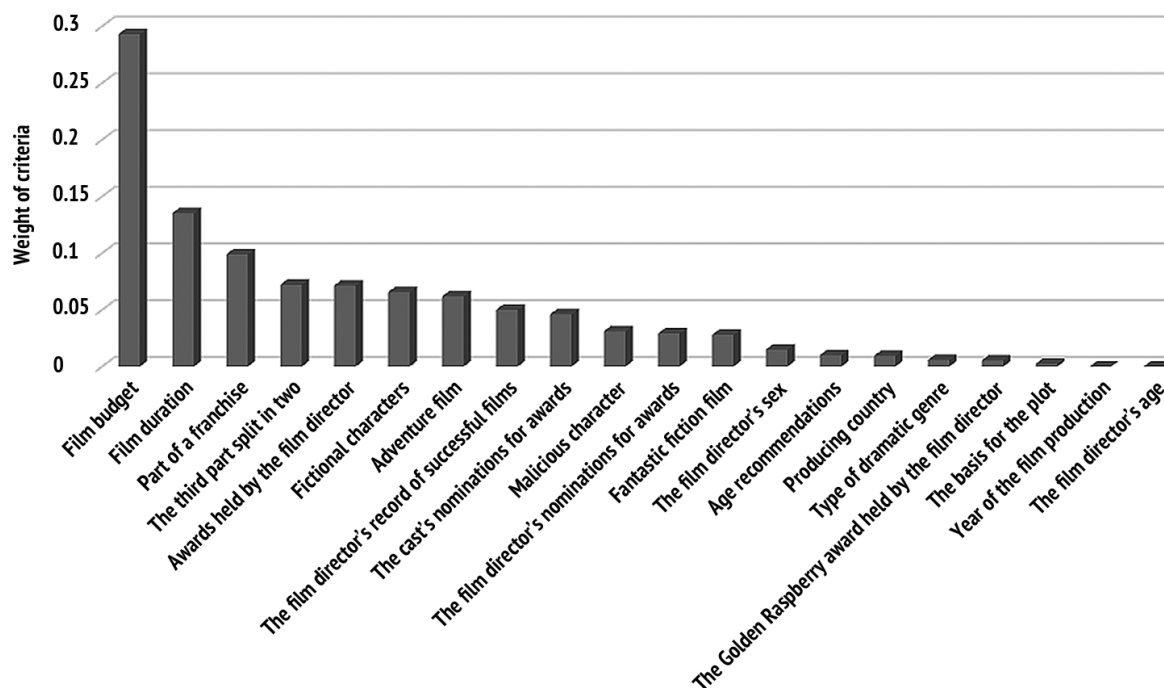
Source: Authoring

Figure 3

Dependence of box-office grosses of *The Da Vinci Code* on its budget, million USD
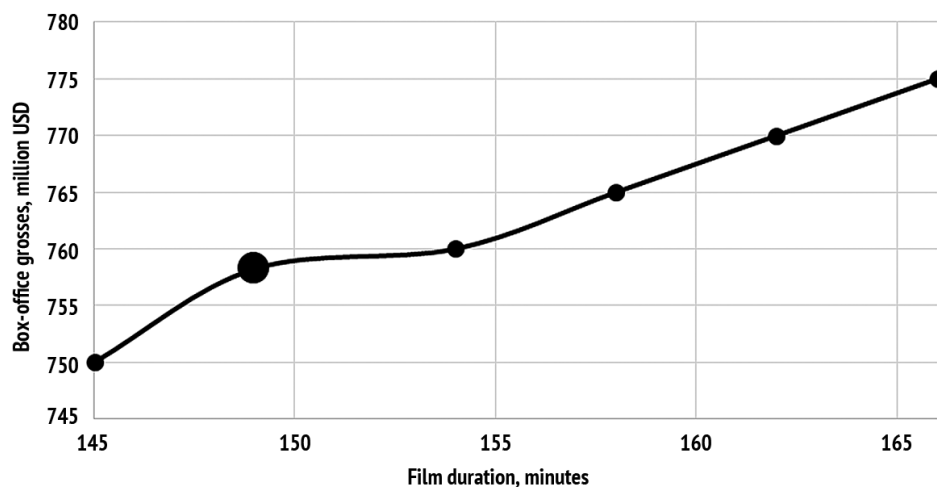

Source: Authoring

Figure 4

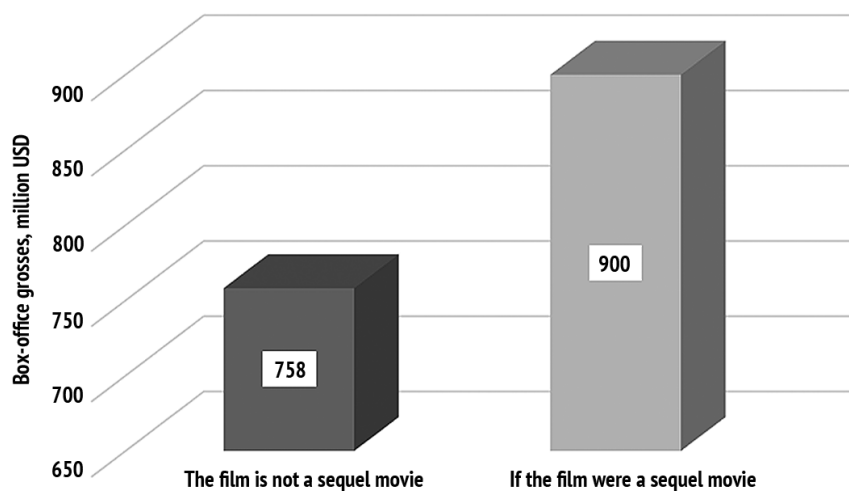**The effect of the budget of** *The Da Vinci Code* **on its profit, million USD**

Figure 5
The effect of film duration on box-office grosses of *The Da Vinci Code*
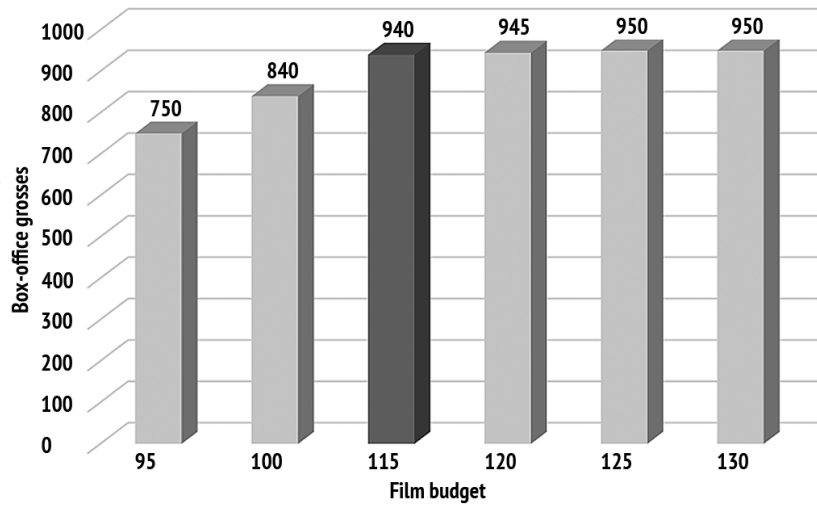

Source: Authoring

Figure 6
The effect of the 'franchise part' aspect on box-office grosses of *The Da Vinci Code*
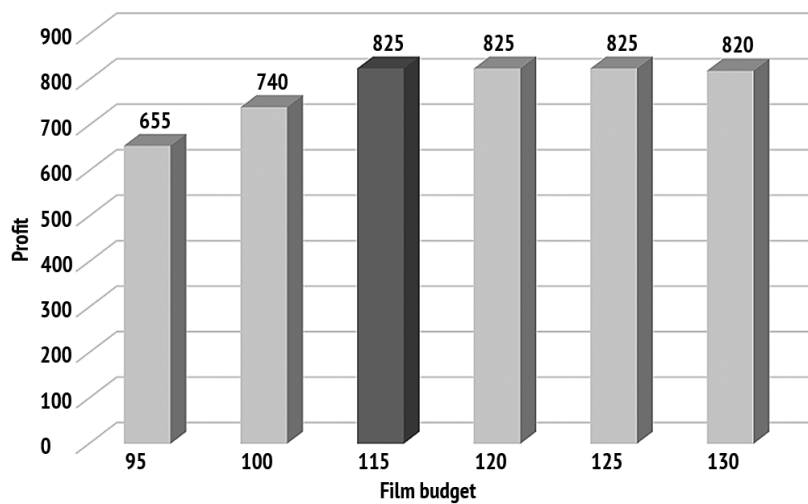

Source: Authoring

Figure 7
Dependence of box-office grosses of *Star Wars. Episode 1: The Panthom Menace* on its budget, million USD

Figure 8
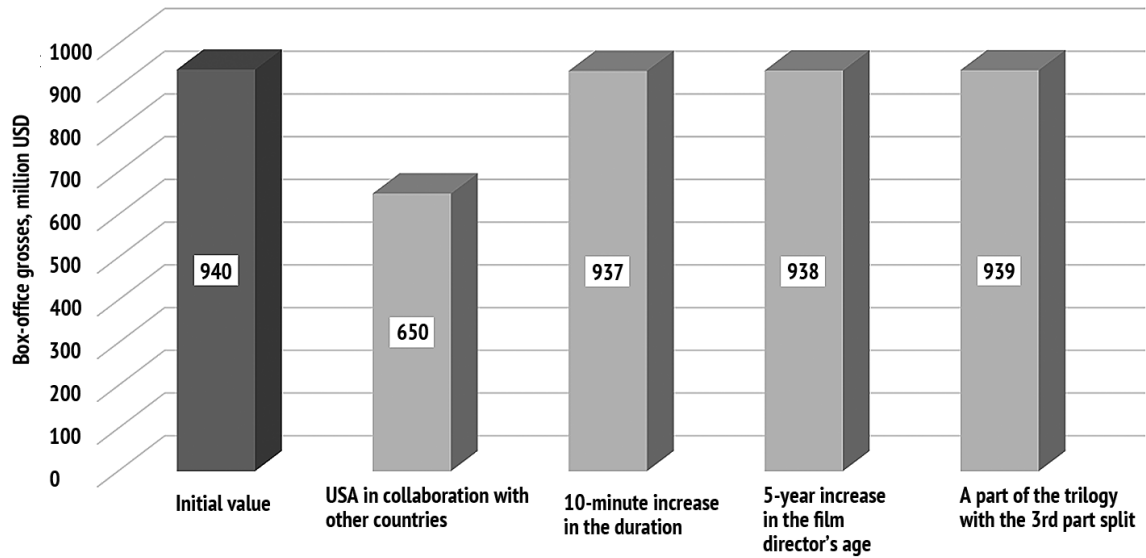**Dependence of box-office grosses on the profit of *Star Wars. Episode 1: The Phantom Menace*, million USD**

Figure 9

The effect of some indicators of *Star Wars. Episode 1: The Phantom Menace* on box-office grosses

Source: Authoring

References

1. Holbrook M.B., Hirschman E.C. The Experiential Aspects of Consumption: Consumer Fantasies, Feelings and Fun. *Journal of Consumer Research*, 1982, vol. 9, iss. 2, pp. 132–140.

2. Eliashberg J., Sawhney M.S. Modeling Goes to Hollywood: Predicting Individual Differences in Movie Enjoyment. *Management Science*, 1994, vol. 40, iss. 9, pp. 1151–1173. URL: https://doi.org/10.1287/mnsc.40.9.1151

3. Sharda R., Delen D. Predicting Box-Office Success of Motion Pictures with Neural Networks. *Expert Systems with Applications*, 2006, vol. 30, iss. 2, pp. 243–254. URL: https://doi.org/10.1016/j.eswa.2005.07.018

4. Litman B.R. Predicting Success of Theatrical Movies: An Empirical Study. *The Journal of Popular Culture*, 1983, vol. 16, iss. 4, pp. 159–175. URL: https://doi.org/10.1111/j.0022-3840.1983.1604_159.x

5. Riwinoto M.T., Selly Artaty Zega, Gia Irlanda. Predicting Animated Film of Box-Office Success with Neural Networks. *Jurnal Teknologi*, 2015, vol. 77, iss. 23, pp. 77–82.

6. Nevolin I.V., Tatarnikov A.S. [Models to project box-office grosses of film-making on the basis of emotional drivers of demand]. *Ekonomika i sotsium*, 2014, no. 4, pp. 1244–1259. (In Russ.) URL: http://iupr.ru/domains_data/files/sborniki_jurnal/Zhurnal%20_4(13)%202014%204.pdf

7. Wasserman M., Mukherjee S., Scott K. et al. Correlations Between User Voting Data, Budget and Boxoffice for Films in the Internet Movie Database. *Journal of the Association for Information Science and Technology*, 2015, vol. 66, iss. 4, pp. 858–868. URL: https://doi.org/10.1002/asi.23213

8. Ghiassi M., Lio D., Moon B. Pre-Production Forecasting of Movie Revenues with a Dynamic Artificial Neural Network. *Expert Systems with Applications*, 2015, vol. 42, iss. 6, pp. 3176–3193. URL: https://doi.org/10.1016/j.eswa.2014.11.022

9. Dhar T., Sun G., Weinberg C.B. The Long-Term Box Office Performance of Sequel Movies. *Marketing Letters*, 2012, vol. 23, iss. 1, pp. 13–29. URL: https://doi.org/10.1007/s11002-011-9146-1

10. McCulloch W.S., Pitts W.A. Logical Calculus of Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, 1990, vol. 52, iss. 1-2, pp. 73–97.

11. Cherepanov F.M., Yasnitskii L.N. [Neural network filter for excluding outliers in statistical data]. *Vestnik Permskogo universiteta. Seriya: Matematika. Mekhanika. Informatika = Perm University Herald. Series: Mathematics. Mechanics. Informatics*, 2008, no. 4, pp. 151–155. (In Russ.)

12. Rosenblatt F. Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. New York, Spartan Books, 1962, pp. 245–248.

Conflict-of-interest notification

We, the authors of this article, bindingly and explicitly declare of the partial and total lack of actual or potential conflict of interest with any other third party whatsoever, which may arise as a result of the publication of this article. This statement relates to the study, data collection and interpretation, writing and preparation of the article, and the decision to submit the manuscript for publication.