

О.В. Волченко, В.И. Корсунова

*Национальный исследовательский университет «Высшая школа
экономики», г. Санкт-Петербург*

АНАЛИЗ INTERNET MOVIE DATABASE: СОЦИОЛОГИЧЕСКИЕ ПЕРСПЕКТИВЫ

Тема культурного производства затрагивается во множестве исследований. Большинство из них сфокусировано на смыслах и символах произведения и выполнено в качественной перспективе [5]. В них биографии и личная жизнь авторов рассматривается как основа создания того или иного культурного продукта [4; 8]. В других работах уделяется внимание экономическим аспектам культурного производства [2]. Помимо этого, некоторые исследования используют стратегию кейс-стади и описывают характеристики определенного фильма для иллюстрации паттернов кинопроизводства [1].

В данном докладе предполагается использование Интернет-базы кинофильмов (Internet Movie Database, сокращенно IMDb), являющейся онлайн-ресурсом, содержащим информацию обо всех созданных фильмах и телепрограммах. IMDb включает данные о ключевых словах, годе и стране производства фильма, жанре, именах режиссеров, сценаристов, артистов, составе съемочной группы и т.д. На данный момент в базе содержится информация о более чем 3 млн фильмов. IMDb не предоставляет открытого API-доступа, однако данные могут быть скачаны в текстовом формате.

Несмотря на возможность использования IMDb в социологических исследованиях, на данный момент база чаще используется в компьютерных науках [6], когнитивных исследованиях [7], а также сетевом моделировании [3]. Характерно, что использующие IMDb ученые скорее фокусируются на применении и развитии сложных статистических методов, чем на получении содержательных выводов. В противовес этому мы предполагаем, что IMDb может представлять собой источник социологических данных, отражающих изменения в современном обществе. Анализ IMDb зачастую включает использование других баз данных: netflix, movielens и др. Тем

не менее, исследователи не принимают в расчет возможность совместить данные о фильмах с социальными характеристиками. Мы предполагаем, что исследование кинопроизводства как интегрированного в социальные процессы феномена, требует включения в анализ других социальных показателей (European Values Study, European Social Survey, World Values Survey, данные Всемирного Банка и др.).

Помимо совмещения IMDb с другими базами данных, мы рассматриваем возможность применения различных статистических процедур для получения ценных для социологов выводов. Например, сетевой анализ ключевых слов может быть использован для выявления связанных семантических полей в кинематографе. Регрессионное моделирование применимо для объяснения появления фильмов определенной тематики в каждой отдельной стране в зависимости от характеристик этой страны. Кроме того, для изучения социальных аспектов развития киноиндустрии и кросс-временной вариативности может быть использован анализ временных рядов.

Литература

1. *Chopra-Gant M.* Hollywood genres and postwar America: masculinity, family and nation in popular movies and film noir. London: IB Tauris, 2005.
2. *De Vany A., Walls W.D.* Does Hollywood Make Too Many Rated Movies? Risk, Stochastic Dominance, and the Illusion of Expectation // *The Journal of Business*. 2002. Vol. 75. № 3. P. 425–451.
3. *Fatemi M., Tokarchuk L.* An empirical study on IMDb and its communities based on the network of co-reviewers // *Proceedings of the First Workshop on Measurement, Privacy, and Mobility*. N.Y.: ACM, 2012. P. 7–10.
4. *Hauser A.* *The Social History of Art*. Vol. 4: Naturalism, impressionism, the film age. Psychology Press, 1999.
5. *Laderman D.* *Driving visions: exploring the road movie*. University of Texas Press, 2010.
6. *Peralta V.* *Extraction and integration of movielens and imdb data*. Technical Report, Laboratoire PRiSM, Université de Versailles, France, 2007.
7. *Rubin T., Steyvers M.* A topic model for movie choices and ratings // *Proceedings of the 9th International Conference on Cognitive Modeling*. 2009.
8. *Zolberg V.L.* *Constructing a Sociology of the Arts*. Cambridge University Press, 1990.