

# Analysis of the film selection process for a film festival

Masashi Inoue  
Graduate School of Science and Engineering  
Yamagata University  
Yonezawa, Yamagata, Japan 992-8510  
Email: mi@yz.yamagata-u.ac.jp

Sakie Sakuma  
Department of Engineering  
Yamagata University  
Yonezawa, Yamagata, Japan 992-8510

## Abstract

This paper presents the results of our analysis of the film selection process for a film festival. The selection of films characterises the festival, which then attracts audiences. Although the criteria for selection are fundamentally decided by members of the selection committee, we investigate if computation can characterise the selection process by utilising available data in this study. Selections are made by watching films and not by checking the associated information. However, considering the difficulties in semantically understanding the video content for computers, we attempted the analysis of the selection process based solely on associated information, including textual, numerical, or categorical data. Results of the analysis revealed factors that may affect the outcome of the selection.

## 1. Introduction

Films have been a major form of entertainment since their appearance in the late 1880s. Cinemas or cinema-like places were originally the only places where films could be viewed; however, copying and projection costs have decreased, and films can now be watched in a number of formats and locations. The recent explosion of digital media has changed the way people consume video material, and films can even be viewed on hand-held devices. Many films are consumed on a demand basis. In such circumstances, film festivals that display films live play a unique role for film viewers.

By using computing analysis to understand the attraction of specific film festivals, two aspects need

to be considered. The first aspect is related to the films themselves: the films are shown in the event. The second aspect is related to the actual event: the organisation of the programme. This paper focuses on the first aspect. This paper is organised as follows. First, we introduce the film festival including its film selection process. Second, we explain the details of the data and variables used. Third, the analysis results are introduced. Finally, the related work is introduced and the paper is concluded.

## 2. Yamagata International Documentary Film Festival (YIDFF)

### 2.1 Overview

The character of film festivals is heavily dependent on the exhibited films. Films are selected by programme committees, and the selection criteria vary among festivals. In this paper, we analysed the Yamagata International Documentary Film Festival (YIDFF)<sup>1</sup>. YIDFF is one of the most recognised documentary film festivals in Asia. YIDFF is renowned worldwide for its artistic rather than commercial selection of films and intimate atmosphere. The organisation of YIDFF is also compared with other Japanese non-documentary film festivals [1]. YIDFF has been held in October in Yamagata Japan biannually since 1989. Five sites were used in Yamagata City, and in 2011, the festival was eight days long. Although several events are scheduled during the event, such as

1. <http://www.yidff.jp/>

workshops and symposia, the primary content of the festival is film screening. The main screening programme consists of the two competition programmes: International Competition (IC) and New Asian Currents (NAC).

## 2.2.Film Selection

IC accepts submissions of full-length films from all over the world, and about 15 films are selected for inclusion in the programme regardless of the number of applications. The IC selection committee consists of YIDFF office judges as well as external judges. For each film, at least two judges are assigned. In the first step, each judge watches the assigned films (about 300 films) from beginning to the end without fast-forwarding and evaluates them in three grades with comments. Secondly, based on the evaluation of judges, the committee selects films after discussions and consultations. This process may include additional viewing by committee members other than the initial judges. About 10 selection advisers support the selection committee but do not participate in the evaluation and selection.

NAC introduces emerging filmmakers from across Asia. Directors must be either residing in or originally from an Asian country. YIDFF office members select the films for this competition. The process of selection is the same as that for IC except for the composition of the committee.

The number of applications is on the increase: 1,218 in 2001 to 1,783 in 2011 as shown in Figure 1. As the number of selected films to the IC and NAC programmes has not increased, the film festival has become more competitive for filmmakers.

During the film festival, international juries are invited, and they evaluate the selected films in the IC programme. The best work is given the Robert and Frances Flaherty Prize. Other awards include the Mayors Prize, two awards of excellence, and Special Prize. For the NAC programme, the most promising work is awarded the Ogawa Shinsuke Prize. Two awards of excellence are likewise given out. In addition to the above awards for each programme, the following programme-wide awards are conferred on deserving works. The Citizens Prize is based on the number of votes by audiences during the festival. The Community Cinema award is given by the Japan Community Cinema Centre.

The jury is composed of people involved with screening films in these localities. The Directors' Guild of Japan Award is given to a film in the NAC and Perspectives Japan programmes that portrays the triumph of a people over various difficulties faced by their nation. The Sky Perfect IDEHA Prize is given to the most promising director of the Japanese films shown in the IC, NAC, and Perspectives Japan programmes. In addition to the programmes, these awards also characterise the film festival.

## 3. Data

### 3.1.Data Content

We used the application data of 2007, 2009, and 2011. Although 1,633, 1,796, and 1,783 applications were submitted for these years, available data only cover 1,361, 1,653, and 1,501 applications, respectively. Missing films from the database include one NAC film in 2007, one IC and four NAC films in 2009, and eight NAC films in 2011.

### 3.2.Target Variable

We consider two target variables: selected/unselected and awarded/un-awarded. 'Selected' indicates that a submitted film was chosen to be screened during the festival. We consider the final decisions made by the selection committee but not the evaluations made by individual committee members. 'Awarded' means that a screened film was conferred with one or more of the available prizes. A single film may be awarded multiple prizes; in this case we do not take the number of prizes into consideration. These two variables are both binary. In the review of films, different elements are considered, including the descriptions of films and their evaluation [3]. However, as we do not know how the films were recommended by the selection committee or the juries, we use the binary values of films (selected/unselected or awarded/unawarded).

### 3.3.Independent Variables

The data consist of the following elements: name of film, production country, production year,

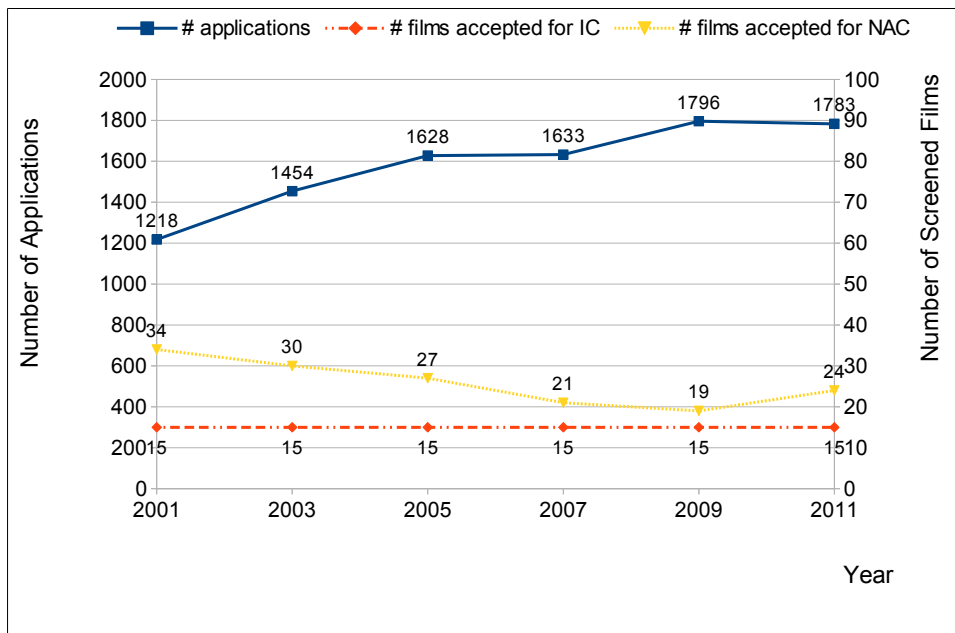


Figure 1. Number of applications and accepted films for each programme from 2001 to 2011.

language, colour (black and white or colour), film format (print or video), running time, name of director, name of producer, name of scriptwriter, name of camera person, name of editor, name of sound recorder, name of music person, name of narrator, name of other staff, name of production company, name of distributor, synopsis, and applied competition programme. These pieces of information were provided by the applicants by filling out the entry form, and are accessible to all through the YIDFF database with the web search interface except for the data on applied programme and application year, regardless of the film selection status<sup>2</sup>. The database contains the information of all submitted films since 1997, excluding those that the applicants declined to be included. Some information may be missing in some films because of the deficiency of the original entry forms. Regarding the titles and synopses, when the entry forms were only filled either in English or in Japanese for

2. <http://www.yidff.jp/search/search-e.html>

some films, the YIDFF office might provide their translation for the missing language. We firstly analysed the correlation between these variables and target variables. However, the relations found were not clearly interpretable. We then focused on the following four factors.

## 4. Analysis

### 4.1 Sentiment Analysis

Among the available metadata, the synopsis is the most informative on the content of films. We were interested in the influence of the mood of films in the selection results. Sentiment analysis has been used for determining the mood of movie reviews [4], and we expect that it can be applied to film synopses as well. Therefore, we conducted sentiment analysis on the Japanese synopses to estimate the positivity of documents based on the

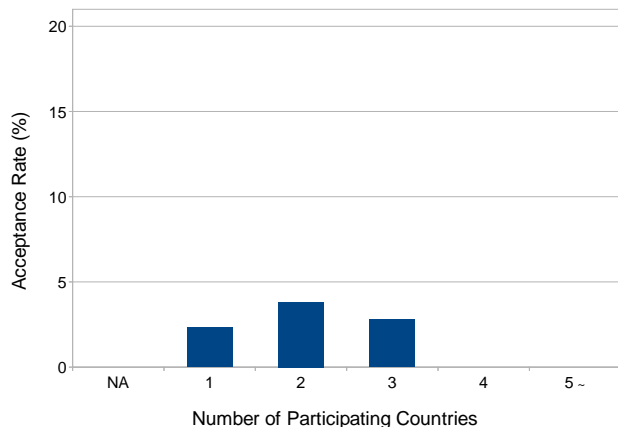


Figure 2. Acceptance ratio of submitted films with different numbers of countries involved (YIDFF 2007).

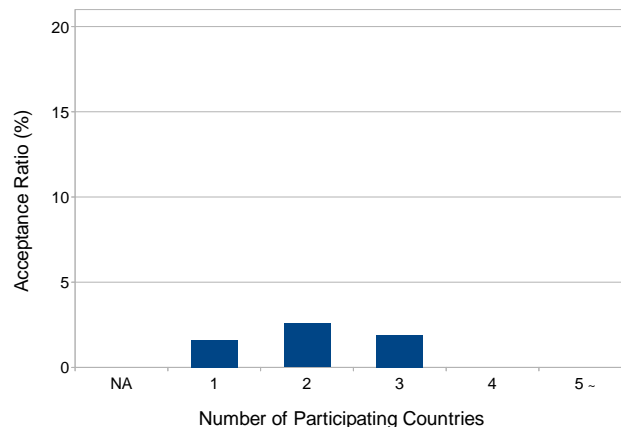


Figure 3. Acceptance ratio of submitted films with different numbers of countries involved (YIDFF 2009).

dictionary of word sentiments [5]. Each word in the dictionary has a positivity score ranging from  $-1$  to  $+1$ . Each word in a synopsis added its score. The accumulated scores were then divided by the number of words in the synopsis for the purpose of normalisation. Results of the calculation showed insignificant correlation between the sentiment scores and the selection or award results. This finding may be attributed to the limitation of the sentiment dictionary. We manually inspected the calculated scores and found that the scores did not accurately indicate the mood of the synopses: synopses with high scores did not give positive impressions, whereas those with low scores did not give negative impressions. We found that the dictionary consists of about 90% of negative words and 10% of positive words. Moreover, it contains many compound words, which skewed the analysis results given that the scores were dependent on the amount of compound words in the synopses. Therefore, sentiment analysis is likely incapable of estimating the actual mood of the synopses with current setting.

#### 4.2. Collaboration Analysis

Modern documentary films are often produced on a multinational level, and we observed that the recognised films often involved people from multiple countries. As the application data contain

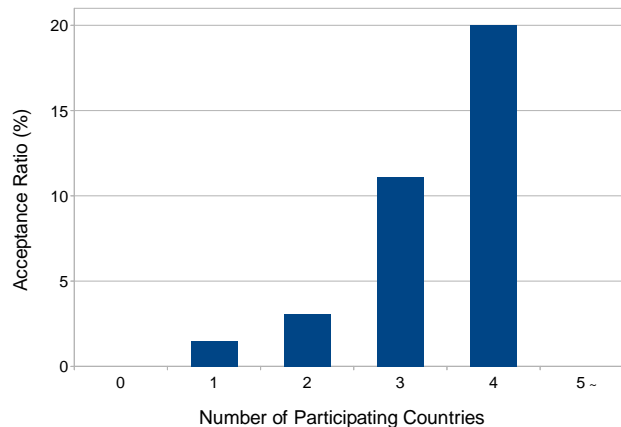


Figure 4. Acceptance ratio of submitted films with different numbers of countries involved (YIDFF 2011).

information on the countries involved, we analysed the effects of international collaboration in terms of the number of participating countries in the films. Figures 2, 3, and 4 show the relationship between the film acceptance rate and the number of countries involved in the films. NA indicates missing data. As shown in Figure 4, a film that involved more people from many countries has a higher likelihood of being selected; in 2011, films

Table 1. Variations in running time format using 11 minutes, 22 seconds, and 33 milliseconds as example.

Time Format	Ratio (%)
11	96.8
11 (recording time $x$ minutes)	2.1
11:22:33	1.0
Others (11'22, 11/22, 11.22)	0.1

with four countries involved had a 20% acceptance rate, which was significantly larger than the average acceptance rate of all applied films. This tendency was not observed in 2007 and 2009. This result may indicate the importance of international collaboration in film making. We will observe if this trend continues in future festivals.

International collaborations come in different forms. A typical example is when an Asian directors receives funding from European or Japanese funding agencies. Another example is when a director immigrates to a Western country, and visits the original country to shoot there with the staff from the present country. The reason for the higher acceptance rate for internationally collaborated films might be different depending on the different collaboration styles. However, we can expect that the film quality benefits from being well funded or being helmed by directors trained in developed countries.

### 4.3. Time Format Analysis

To compare the running times of applied films, we standardised the time format. We extracted raw time values from the time information containing various text and punctuation marks. Then, we rounded up the number into minutes. Although the running time did not influence selection, we noticed variations in the submitted entry forms, especially in the format of running time during data cleansing. Examples of time formats are shown in Table 1.

The submission format has no restriction, but most applicants used the first format that uses minutes only. Selection committee members do not care about the format; however, almost all selected films used the standard format as shown in Table 1. We suspect that the preference for the standard

Table 2. Ratio of films with distributors (%).

	2007	2009	2011
Applied to IC	53.6	50	48.1
Selected to IC	53.3	78.6	71.4
Awarded in IC	33.3	80	60
Applied to NAC	23.2	24.7	20.5
Selected to NAC	26.3	46.7	29.4
Awarded in NAC	0	25	28.6

format may reflect the importance of applicants experience. If a filmmaker, or a producer who supported the application, has more experience in the submission of films to film festivals, he or she may use the typical format in writing the running time. The time format may reflect the experience of applicants and indirectly influence the selection results. Although we used the example of time format in this work, other traits may directly or indirectly indicate the expertise or preparedness of applicants.

### 4.4. Pre-Show Attention Analysis

Some films are ballyhooed, whereas other films are totally unknown prior to their application to the festival. Films with better reputation may have a higher chance of acceptance because their quality is considered higher even if the selection committee members are not aware of them. We took the distributor information as the indicator of prior reputation of the films. If a film already has a distributor, it has likely been judged as attractive. Table 2 presents the results of our analysis. As for IC, in 2009 and in 2011, the ratios of selected films with distributors were 78.6% and 71.4%, higher than those selected film without distributors at 50% and 48.1%, respectively. These numbers clearly indicate that the existence of distributors prior to the application reflects film quality. In contrast, for NAC, although the ratio of selected films with distributors increased from 24.7% applied to 46.7% selected, the existence of distributors was not a clear indication of influence. The difference may be explained by the characteristics of the two programmes: IC accepts high quality films from around the world, whereas NAC intends to introduce up-and-coming Asian documentary filmmakers.<sup>3</sup>

## 5. Related Work

Our work is similar to film recommendation in that it selects relevant films among many candidates based on stated film characteristics [6]. Our work focuses on the understanding of the film selection process by experts. If the process is sufficiently understood, a recommendation model can be built based on accurate knowledge. In building a film recommendation system for film selection in film festivals, a key difference is that the system is intended for film experts and not consumers.

## 6. Conclusion

We analysed the film selection process in the YIDFF from 2007 to 2011. Selections were performed by human experts based on film content. In contrast, we attempted to analyse computationally the related data of candidate films to shed light on the selection process. We aimed to identify features from the metadata that may predict selection results. Although the relationships still need to be investigated using future festival data, we found interesting relations between the selection and the number of countries involved, format of entry form, and existence of distributors. However, we could not find a single strong indicator for the selection. This result supports the premise that the selection committee members carefully investigate the content of candidate films and eliminate the influence of contextual information.

Factors that we have not fully examined are as follows: the language used in films, nationalities of directors, existence of film distributors at the time of application and others. The influence of these factors in the selection results would be an interesting future research topic. Another interesting direction is to estimate selection biases in the YIDFF selection committees. For example, Kuei-fen Chiu pointed out that Taiwanese documentary films awarded at YIDFF are mostly psychological in tone rather than social because of the lack of geopolitical knowledge of the selection committee [7]. Such biases brought about by the

3. The number of applications is different from the number announced by the YIDFF office. We used the raw information used for the internal data organization before the data cleansing.

factor of internationality are not derivable from the currently available data. We have to explore alternative methods to analyse the selection process in further detail.

## Acknowledgment

This research was partially supported by the Grant-in-Aid for Scientific Research 24500321. The authors would like to thank Katsuyoshi Kusakabe and Hitomi Ochiai for the information on the YIDFF.

## References

- [1] L. MacGregor and E. Nickanen, "Contemporary film festivals in japan : Part1," *Gengo, Bunka, Shakai*, vol. 11, pp. 75–113, Mar 2013.
- [2] L. Manzor, K. Rimkus, and M. Ogihara, "Cuban theater digital archive: A multimodal platform for theater documentation and research," *ECLAP 2013*, vol. LNCS 7990, pp. 138–150.
- [3] M. Taboada, "Stages in an online review genre," *Text & Talk*, vol. 31, no. 2, pp. 247–269, 2011.
- [4] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: Sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, ser. EMNLP '02, Stroudsburg, PA, USA, 2002, pp. 79–86.
- [5] H. Takamura, T. Inui, and M. Okumura, "Extracting semantic orientations of words using spin model," in *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, 2005, pp. 133–140.
- [6] N. Koenigstein and U. Paquet, "Xbox movies recommendations: Variational bayes matrix factorization with embedded feature selection," in *Proceedings of the 7th ACM Conference on Recommender Systems*, 2013, pp. 129–136.
- [7] K.-F. Chiu, "Afterword: Documentary filmmaking as ethical production of truth," *Concentric: Literary and Cultural Studies*, vol. 39, no. 1, pp. 203–220, March 2013.