Movie Recommendation Systems Using Actor-Based Matrix Computations in South Korea

Syjung Hwang and Eunil Park

***Abstract*— A recommendation system saves the user the trouble of searching for information and analyzes their profile to recommend the most suitable content. A variety of techniques, including content-based, collaborative, and knowledge-based techniques, have been proposed for performing recommendations. Recommendation systems are employed to suggest content, such as books, music, and videos; furthermore, they are often used in e-commerce. In particular, the South Korean film industry recommends movies using a collaborative filtering method based on genres, which is commonly used in film recommendation systems. However, this method can be ineffective when users initially encounter film recommendation services or have specific preferences with respect to movies, such as preferences regarding actors or directors. This motivated us to propose an actor-based recommendation system using the content-based filtering that considers actors' filmography information and the genres of 509 South Korean movies. The effectiveness and performance of the suggested system are evaluated in comparison with those of a traditional genre-oriented recommendation system.**

***Index Terms*— Actor, content-based recommendation, movie recommendation system, South Korea.**

I. Introduction

RAPID advances in Internet and data technologies have led to the development of a number of recommendation systems to suggest various contents to users. This is convenient because users do not need to search numerous websites to find the contents that are of interest to them. Because the contents have many unique characteristics and features, scholars and organizations have explored the use of recommendation systems and solutions to achieve success in competitive markets [1]–[3].

Considering this trend, several notable recommendation approaches, including content-based, collaborative, knowledge-oriented, and hybrid techniques, have been introduced [4], [5].

Collaborative filtering techniques focus on the fact that users who have similar characteristics tend to have similar preferences [6]. Compared to collaborative filtering techniques, content-based filtering techniques consider previous evaluations performed by users to predict their next selections [7]. These two representative recommendation techniques have been used by many companies in their user-customized services [8].

Recommendation systems have become indispensable for most online services, including bookstores and video stores [9]. Among such services, online movie streaming platforms, such as Netflix [10] and YouTube [11], provide user-customized video contents through their own platforms. Based on the massive amount of data collected from users on these platforms, advanced collaborative filtering-based techniques are employed [12]. Although both services have been exceedingly successful in the marketplace, several notable limitations have hampered the use of these techniques by these platforms. Other techniques, including content-based filtering techniques, should be employed by platforms that do not have a sufficient number of users; the latter technique should also be employed when specific users initially encounter the services. This means that the use of collaborative filtering techniques in the initial stage of these services could be problematic [13], [14].

Based on the above-mentioned difficulties, several scholars working on movie recommendation systems attempted to employ collaborative or hybrid filtering recommendation systems with particular categorized standards, such as movie genres [15], [16]. For instance, Netflix also operates its recommendation system by reflecting individual users' propensity for each genre [17]. Although this approach can be helpful and useful for users who aim to select a specific movie by considering movie genres, the approach would be ineffective in the case of users with specific preferences with respect to movies, such as preferences regarding actors or directors.

This study proposes a content-based recommendation system in the form of an actor-based recommendation system that considers actors' filmography information and the genres of 509 South Korean movies. Then, the effectiveness and performance of the suggested system are assessed in comparison with those of a traditional genre-oriented recommendation system.

Manuscript received 3 February 2021; revised 25 August 2021; accepted 21 September 2021. Date of publication 11 October 2021; date of current version 30 September 2022. This work was supported in part by the Ministry of Science and ICT (MSIT), South Korea, through the ICT Challenge and Advanced Network of HRD (ICAN) Program supervised by the Institute of Information and Communications Technology Planning and Evaluation (IITP) under Grant IITP-2020-0-01816 and in part by the IITP Grant through the Korean Government (MSIT) (AI Big Data Based Cyber Security Orchestration and Automated Response Technology Development) under Grant 2021-0-00358. *(Corresponding author: Eunil Park.)*

Syjung Hwang is with the Department of Interaction Science, Sungkyunkwan University, Seoul 03063, Republic of Korea, and also with Naver Webtoon, Seongnam 13529, Republic of Korea.

Eunil Park is with the Department of Interaction Science, Sungkyunkwan University, Seoul 03063, Republic of Korea (e-mail: eunilpark@skku.edu).

Digital Object Identifier 10.1109/TCSS.2021.3117885

The remainder of this article is organized as follows. First, we provide an overview of content-based recommendation systems and their limitations. Second, we present the new recommendation approach with its results. Finally, notable implications and limitations are examined and future research areas are suggested.

II. Related Work

A. Content-Based Recommendation System

A content-based recommendation system, which mainly relies on cognitive filtering, is "a system that displays item recommendations by comparing alternative items with those associated with user profiles" [18]–[20]. This system mainly analyzes the features of specific items that were previously selected by a specific user and extracts the unique features of these items to recommend other items belonging to similar categories. For instance, a specific user who viewed a number of romance movies in the past would receive suggestions for other romance movies the user has not yet viewed. Systems such as this are already employed by search engines, bookstores, and shopping malls [21].

In general, a content-based recommendation system does not consider similarities between users when recommending specific items. Instead, the system focuses entirely on the characteristics of the contents used by an individual user. Thus, in terms of the individual user, this system is considered to be more reliable than other recommendation systems because it only considers the unique features of the content related to the individual user profile [22].

The main purpose of content-based recommendation systems that take both item attributes and user profiles into account is to suggest more suitable items to users. This means that the system would need to employ methods and classification solutions that are more heuristically oriented [23], [24].

B. Problems With a Content-Based Recommendation System

Although content-based recommendation systems have been widely employed, they have several notable limitations. First, the system could be limited by the nature of content features. To provide well-defined recommendations, the system requires a sufficient number of features [25]. However, it is difficult to extract features from videos or images; thus, these kinds of data cannot provide large numbers of features for content-based recommendation system. Second, because other users' evaluations are not considered when the suggestions are made, it is difficult to guarantee the quality of the suggested contents. This means that a limited number of items can be suggested when the system has notable restrictions in terms of recommending items (e.g., recommending only highly rated items) [1]. This may lead to notable polarization problems when suggesting and recommending specific items [26].

Because of these limitations, content-based recommendation systems are not necessarily able to recommend movies effectively. To address this shortcoming, we employed film industry-oriented perspectives toward content-based movie recommendation systems. Although a number of features are associated with a specific movie, the actors, especially those who are movie stars, have a considerable impact on the box office success of movies. In other words, people tend to choose movies based on their favorite actors; therefore, the development of a recommendation system that considers users' favorite actors would require a correlation indicator of this genre based on actors [27]–[29]. Our study, therefore, aims to develop a content-based recommendation system that considers movie actors in the South Korean film industry.

III. Method

A. Data Collection and Preprocessing

We obtained the data from the Korean Film Council.[1] It included a list of movies, their details, film companies, actors, and directors. We extracted the unique code, title, production year, opening year, type, production statement, producing countries, and genre of each movie. Also, we obtained the directors' names and their filmography; actors' names and their filmography, date of birth, sex, and role; and the name of the production company and its CEO, together with its filmography and company classification. The data included 3429 duplicate movie titles, which we identified by assigning a unique number to each title. The movies in the dataset included feature films, short stories, and omnibuses. We decided to use only feature films that had been released.

In addition, we used data from the Naver movie site,[2] which include the title, year, viewers' ratings and the number of viewers, and Internet users' ratings and the number of Internet users of each movie as well as the ratings provided by journalists and critics. The data also contained the following information about each movie: the characteristics, director, actors, grade, cumulative audience attendance, attendance rate, visitor satisfaction by gender, visitor satisfaction by age, direction, acting, story, quality of visual effects, and background music. We only selected movies that were rated by more than 30 people to ensure the reliability of the ratings. The rating of each movie was based on the viewers' ratings, but in the absence thereof, the Internet users' ratings were used instead. This replacement is considered appropriate because of the high correlation between these two types of ratings (0.82) and their similar distribution patterns.

We collected additional data from the Korea Box Office Information System[3]; the data include the movie ranking, movie name, release date, seat sales rate, seat occupancy rate, number of seats, sales, cumulative sales, attendance, and cumulative audience number. The collection contained data regarding 2346 directors, 15 539 actors, and 6647 movies, which are included movies around the world. Because data pertaining to older films were incomplete compared to those regarding recent films, we only used film data from 2010 to 2019. Furthermore, actors and directors of films with an average audience of less than 25 percent were excluded. After preprocessing the data, the dataset included 4450 South Korean actors and 509 South Korean movies (Fig. 1).

[1] http://www.kobis.or.kr/kobisopenapi/homepg/main/main.do
[2] https://movie.naver.com/
[3] http://www.kobis.or.kr/

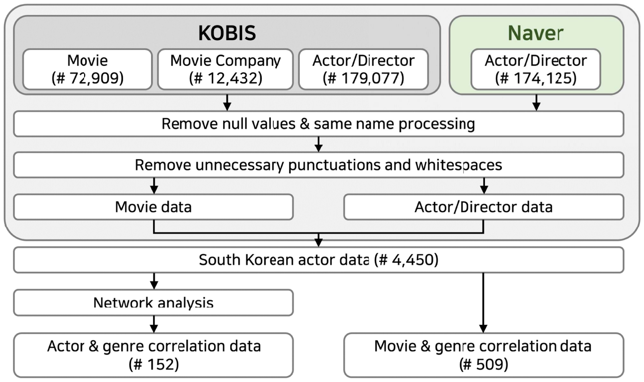

Fig. 1. Data preprocessing procedure for correlation analysis.

TABLE I
GENRE CATEGORY

No.	Genre	No.	Genre
1	Comedy	9	Mystery
2	Fantasy	10	Melodrama
3	War	11	Drama
4	Adventure	12	Horror
5	Action	13	Family
6	Thriller	14	Documentary
7	Historical drama	15	SF
8	Crime	16	miscellaneous

With respect to the movie genres, we excluded the "animation genre," "performance genre," and "musical genre" and combined the "western theater genre" with the "miscellaneous" categories. Table I presents the movie genres used in this study.

Besides, 509 data plots were extracted from the Naver series[4] to implement the content-based recommendation system. The data preprocessing procedure for the data plots we collected is presented in Fig. 2. After collecting the data, we removed unnecessary punctuation marks and white spaces from each plot. Then, we employed Korean natural language processing in python (konlpy), which is a Python package, to tokenize users' posts.

We created the dataset for the content-based recommendation system by combining the actor column, director column, genre column, and plot column to form a string-type column as the content column. Moreover, when recommending movies to users, we added data regarding the viewers' rating, sales, attendance, and sales rate to the final dataset to consider the elements of box office success.

B. Computing the Rank Correlation Between Movie and Genre

Calculating the rank correlation between a specific movie and its genre is based on the combination of genres in the movie database. Each movie has a genre combination, which comprises more than one genre. The movie genres are selected by the movie producers. For example, the movie *26 years* belongs to the genres of "action" and "drama." This means that the movie is categorized according to both of these genres.

[4]https://serieson.naver.com/movie/home.nhn

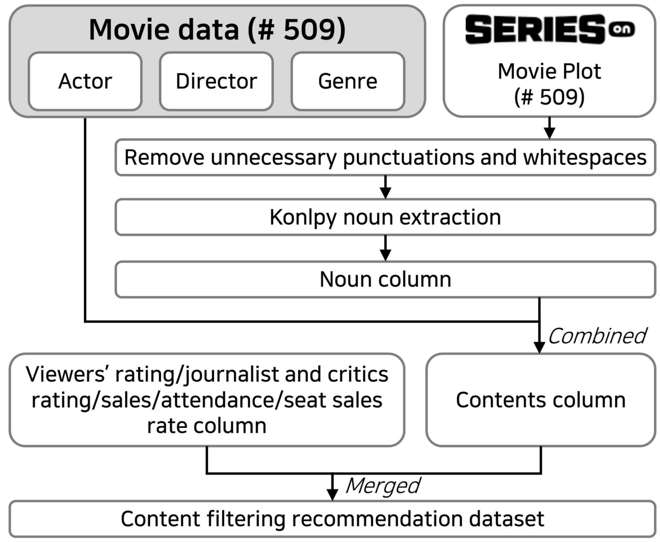

Fig. 2. Data preprocessing procedure for the content-based recommendation system.

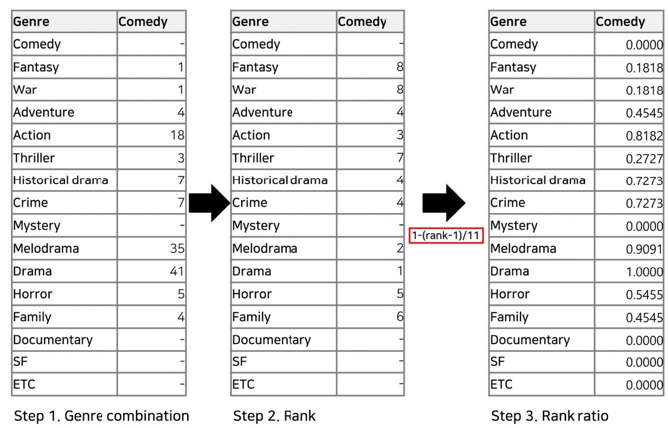

Fig. 3. Course that calculates rank correlations.

To determine the rank correlation between a movie and its genre, we first chose a specific movie and added another between the movie genres. However, a substantial inconsistency exists with regard to the number of movies that belong to each genre. For instance, the total number of movies in the "comedy" genre is 126, and the number of movies in the "fantasy" genre is 12. Thus, rather than simply estimating a correction, we calculated the rank of each genre in each genre column to exclude biased distributions.

Fig. 3 shows an overview of the computational procedures for rank correlation. Table II shows that the values in red differ from those in blue, and the rank correlation between movie and genre was finally calculated by averaging these two parts. For example, the rank correction of "comedy" and "fantasy" in red is 0.33, whereas that in blue is 0.09. These values were averaged to determine the rank correction of "comedy" and "fantasy" to be 0.21.

TABLE II
INCONSISTENCY BETWEEN THE UPPER SIDE (RED) AND LOWER SIDE (BLUE)

	Comedy	Fantasy	War	Adventure	Action	Thriller	Historical drama	Crime	Mystery	Melodrama	Drama	Horror	Family	Documentary	SF	ETC
Comedy	0.00	0.33	0.50	0.75	0.77	0.25	0.40	0.43	0.00	0.91	0.92	0.75	0.60	0.00	0.00	0.00
Fantasy	0.09	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.29	0.64	0.23	0.38	0.00	0.00	0.00	0.00
War	0.09	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.36	0.38	0.00	0.00	0.00	0.00	0.00
Adventure	0.36	0.00	0.00	0.00	0.54	0.00	0.20	0.00	0.00	0.00	0.15	0.00	0.00	0.00	0.00	0.00
Action	0.73	0.33	0.50	0.50	0.00	0.75	0.80	0.86	0.43	0.36	0.85	0.25	0.40	0.00	0.80	0.00
Thriller	0.18	0.00	0.00	0.00	0.69	0.00	0.00	0.57	0.86	0.64	0.62	0.88	0.00	0.00	0.60	0.00
Historical drama	0.64	0.00	0.00	0.25	0.62	0.00	0.00	0.00	0.00	0.64	0.54	0.00	0.00	0.00	0.00	0.00
Crime	0.64	0.00	0.00	0.00	0.85	0.63	0.00	0.00	0.29	0.36	0.77	0.00	0.00	0.00	0.00	1.00
Mystery	0.00	0.33	0.00	0.00	0.54	0.63	0.00	0.29	0.00	0.36	0.54	0.75	0.00	0.00	0.00	0.00
Melodrama	0.82	0.67	0.50	0.00	0.31	0.13	0.20	0.29	0.29	0.00	0.77	0.50	0.40	0.00	0.00	0.00
Drama	0.91	0.83	0.75	0.00	0.92	0.88	0.60	0.71	0.71	0.82	0.00	0.25	0.80	0.00	0.60	0.00
Horror	0.45	0.67	0.00	0.00	0.31	0.38	0.00	0.00	0.57	0.73	0.15	0.00	0.00	0.00	0.60	0.00
Family	0.36	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.36	0.31	0.00	0.00	0.00	0.60	0.00
Documentary	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SF	0.00	0.00	0.00	0.00	0.54	0.00	0.00	0.00	0.00	0.00	0.15	0.25	0.40	0.00	0.00	0.00
ETC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.29	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

TABLE III
GENRES OF MOVIES STARRING THE ACTOR "DON LEE"

Genre	Count	Genre	Count
Comedy	4	Mystery	-
Fantasy	2	Melodrama	2
War	-	Drama	13
Adventure	1	Horror	-
Action	10	Family	-
Thriller	6	Documentary	-
Historical drama	1	SF	-
Crime	9	ETC	-

C. Computing the Correlation Between Actor and Genre

The correlation between a specific actor and movie genre is computed by using the genre combination in the actor database. Because each actor tends to appear more frequently in movies of specific genres than others, the actor can be concluded to be strongly connected with the specific genres in which the actor often appears. For example, from 2010 to 2019, the actor "Don Lee" starred in 13 movies in the drama genre; furthermore, he starred in ten action and nine crime movies. This means that he can be associated with all three of these genres (Table III).

The number of actors who appeared in movies belonging to each genre is not uniform. We, therefore, correlated a specific actor with their genres by using a network analysis to select the top ten actors with a high degree centrality in each genre to reduce this bias. We created a table composed of actors (rows) and genres (columns) and calculated the correlation between a specific actor and their genre using Pearson's correlation coefficient

$$\rho_{XY} = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y}. \quad (1)$$

In (1), X is the selected genre and σ_X is the standard deviation of the number of times an actor appeared in a movie of that particular genre. Let Y be the other genre. The value of Pearson's correlation coefficient is between -1 and 1. If Pearson's correlation between X and Y is closer to 1, then the two genres X and Y are similar in terms of the correlation between a specific actor and their genre.

IV. RESULTS

Figs. 4–6 summarize the results. The correlation between a movie and its genre differs markedly from that between an actor and their movie genres. Even though a significant correlation could exist between movies and their genres (movie producers' perspectives), the relationship between movies and actors cannot be correlated (viewers' perspectives). For example, the "comedy" genre is highly correlated (0.56) with the "adventure" genre when considering movies and genres, whereas a low correlation is observed between actors and genres. Tables IV–VI summarize the correlation matrix between movies and genres.

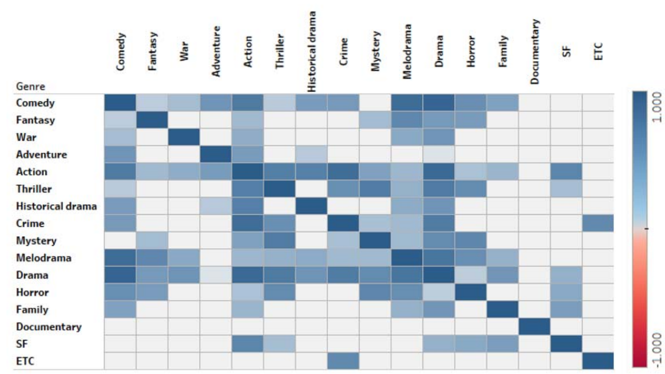

Fig. 4. Rank correlation plot between a movie and its genres.

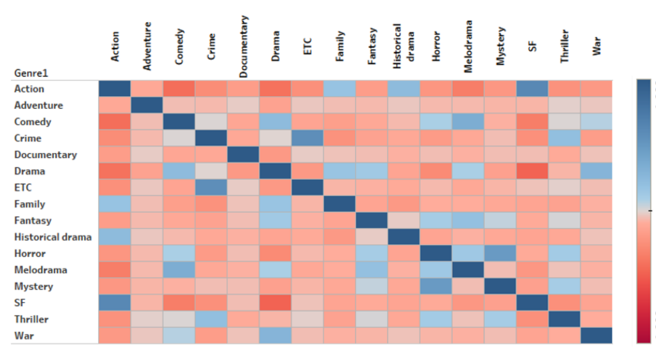

Fig. 5. Correlation plot between an actor and their genres.

We conducted a user survey with ten respondents to evaluate the suggested recommendation system. Five participants responded that they selected their movies by considering the movie genres, whereas the other five participants answered that their movie selections were determined based on the actors featured in the movies.

TABLE IV

GENRES WITH THE RANK CORRELATION GREATER THAN 0.5 BETWEEN A MOVIE AND ITS GENRES, AND GENRES WITH THE CORRELATION OF 0.1 TO −0.1 BETWEEN AN ACTOR AND THEIR GENRES

	Rank correlation between movie and genre	**Correlation between actor and genre**
Comedy	Adventure, Action, Historical drama, Crime, Melodrama, Drama, Horror	War, Adventure, Thriller, Historical drama, Crime, Mystery, Horror
Fantasy	Melodrama, Drama, Horror	War, Adventure, Thriller, Historical drama, Mystery, Document, SF, ETC
War	Drama	Comedy, Fantasy, Adventure, Thriller, Historical drama, Mystery, Document, SF, ETC
Adventure	Comedy, Action	Comedy, Fantasy, War, Thriller, Historical drama, Crime, Mystery, Melodrama, Horror, Family, Document, SF, ETC
Action	Comedy, Adventure, Thriller, Historical drama, Crime, Drama, SF	-
Thriller	Action, Crime, Mystery, Drama, Horror	Comedy, Fantasy, War, Adventure, Melodrama, Drama, Document, ETC
Historical drama	Comedy, Action, Drama	Comedy, Fantasy, War, Adventure, Melodrama, Document, ETC
Crime	Comedy, Action, Thriller, Drama, ETC	Comedy, Adventure, Mystery, Drama
Mystery	Thriller, Drama, Horror	Comedy, Fantasy, War, Adventure, Crime, Melodrama, Document, ETC
Melodrama	Comedy, Fantasy, Drama, Horror	War, Adventure, Thriller, Historical drama, Mystery, Drama, Family, Document
Drama	Comedy, Fantasy, War, Action, Thriller, Historical drama, Crime, Mystery, Melodrama, Family	Thriller, Crime, Melodrama
Horror	Comedy, Fantasy, Thriller, Mystery, Melodrama	Comedy, War, Adventure, Family, Document, SF, ETC
Family	Drama, SF	War, Adventure, Melodrama, Horror, Document, ETC
Documentary	-	Fantasy, War, Adventure, Thriller, Historical drama, Mystery, Melodrama, Horror, Family, SF, ETC
SF	Action, Family	Fantasy, Adventure, Horror, Document, ETC
ETC	Crime	Fantasy, War, Adventure, Thriller, Historical drama, Mystery, Horror, Family, Document, SF

TABLE V

RANK CORRELATION BETWEEN MOVIES AND GENRES

	Comedy	**Fantasy**	**War**	**Adventure**	**Action**	**Thriller**	**Historical drama**	**Crime**	**Mystery**	**Melodrama**	**Drama**	**Horror**	**Family**	**Documentary**	**SF**	**ETC**
Comedy	1.00	0.21	0.30	0.56	0.75	0.22	0.52	0.53	0.00	0.86	0.92	0.60	0.48	0.00	0.00	0.00
Fantasy		1.00	0.00	0.00	0.32	0.00	0.00	0.00	0.31	0.65	0.53	0.52	0.00	0.00	0.00	0.00
War			1.00	0.00	0.40	0.00	0.00	0.00	0.00	0.43	0.57	0.00	0.00	0.00	0.00	0.00
Adventure				1.00	0.52	0.00	0.23	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00
Action					1.00	0.72	0.71	0.85	0.48	0.34	0.88	0.28	0.35	0.00	0.67	0.00
Thriller						1.00	0.00	0.60	0.74	0.38	0.75	0.63	0.00	0.00	0.30	0.00
Historical drama							1.00	0.00	0.00	0.42	0.57	0.00	0.00	0.00	0.00	0.00
Crime								1.00	0.29	0.32	0.74	0.00	0.00	0.00	0.00	0.64
Mystery									1.00	0.32	0.63	0.66	0.00	0.00	0.00	0.00
Melodrama										1.00	0.79	0.61	0.38	0.00	0.00	0.00
Drama											1.00	0.20	0.55	0.00	0.38	0.00
Horror												1.00	0.00	0.00	0.43	0.00
Family													1.00	0.00	0.50	0.00
Documentary														1.00	0.00	0.00
SF															1.00	0.00
ETC																1.00

TABLE VI

GAP BETWEEN THE CORRELATION OF ACTOR–GENRE AND RANK CORRELATION OF MOVIE–GENRE

	Comedy	**Fantasy**	**War**	**Adventure**	**Action**	**Thriller**	**Historical drama**	**Crime**	**Mystery**	**Melodrama**	**Drama**	**Horror**	**Family**	**Documentary**	**SF**	**ETC**
Comedy	0.00	0.10	0.23	0.51	0.33	0.21	0.46	0.52	0.08	0.51	0.68	0.51	0.32	0.12	0.34	0.13
Fantasy		0.00	0.07	0.06	0.15	0.02	0.02	0.14	0.26	0.45	0.40	0.41	0.13	0.05	0.10	0.08
War			0.00	0.03	0.21	0.09	0.04	0.17	0.05	0.35	0.28	0.07	0.08	0.04	0.13	0.05
Adventure				0.00	0.41	0.01	0.20	0.06	0.07	-0.06	0.06	0.06	0.05	0.02	0.07	0.03
Action					0.00	0.49	0.47	0.59	0.28	0.00	0.48	0.07	0.17	0.17	0.04	0.24
Thriller						0.00	0.11	0.39	0.63	0.34	0.68	0.51	0.14	0.10	0.06	0.01
Historical drama							0.00	0.11	0.11	0.34	0.44	0.13	0.19	0.08	0.13	0.10
Crime								0.00	0.24	0.21	0.74	0.18	0.23	0.11	0.24	0.06
Mystery									0.00	0.27	0.48	0.17	0.11	0.05	0.15	0.07
Melodrama										0.00	0.70	0.47	0.29	0.08	0.20	0.11
Drama											0.00	0.07	0.38	0.19	0.09	0.19
Horror												0.00	0.09	0.05	0.33	0.06
Family													0.00	0.04	0.37	0.07
Documentary														0.00	0.05	0.02
SF															0.00	0.04
ETC																0.00

In the survey, once a respondent selected their favorite movie, the system recommended 20 movies that were organized as the ten movies most highly correlated with their genres and the ten movies with the highest correlation between the actors and genres. Then, each participant indicated their preference for the recommended movies on a ten-point Likert scale (1: "strongly not preferred" to 10: "strongly preferred") [30]. For instance, Table VII contains the movie list

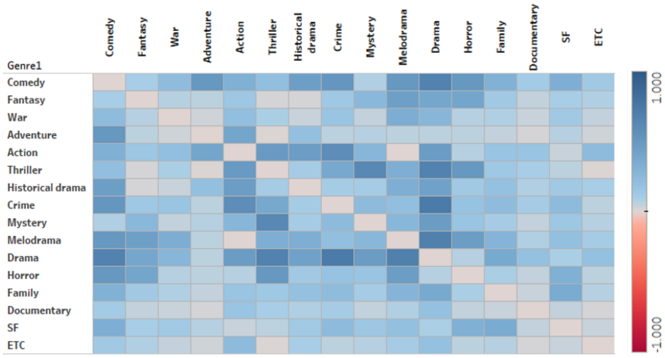

Fig. 6. Difference between the correlation between an actor and their genre and the rank correlation between a movie and its genre.

TABLE VII

RESULTS OF MOVIE RECOMMENDATIONS FOR PARASITE; "MOVIE–GENRE" CORRESPONDS TO THE CASE WHERE THE USER'S MOVIE SELECTION CRITERION IS THE GENRE, WHILE "ACTOR–GENRE" CORRESPONDS TO THE CASE WHERE THE USER'S MOVIE SELECTION CRITERION IS THE ACTOR

Movie - Genre		Actor - Genre	
No.	**Title**	**No.**	**Title**
1	*A Taxi Driver*	1	*A Taxi Driver*
2	*The Attorney*	2	*The Attorney*
3	*Ode to My Father*	3	*Ode to My Father*
4	*Blind*	4	*Blind*
5	*Secret Reunion*	5	*Secret Reunion*
6	*The Neighbors*	6	*The Neighbors*
7	*The Age of Shadows*	7	*i Can Speak*
8	*Snow piercer*	8	*1987*
9	*i Can Speak*	9	*Default*
10	*1987*	10	*All About My Wife*

TABLE VIII

COMPARISON OF TEST RESULTS

User	Preference	Genre-based	Actor-based
1	genre	8	7.5
2	genre	8	8
3	actor	5	6
4	actor	8.5	9.5
5	actor	8	9
6	genre	8	7
7	genre	7	6
8	actor	5	8
9	actor	7.5	9
10	genre	9	9
11	actor	5	6
12	actor	8	9
13	actor	6	6.5
14	actor	7	8
15	genre	8	7
16	genre	9	7
17	actor	6	9
18	actor	7	8.5
19	actor	7.5	8
20	genre	8	7.5
Average	Genre: 8, Actor: 12	7.28 (1.23)	7.78 (1.10)

that was presented to a respondent who selected *Parasite* as their favorite movie.

Table VIII shows the survey results. Users, who select movies based on actors, were more likely to prefer the movies recommended by the actor–genre approaches than those by the genre-oriented approaches ($p < 0.05$).

V. CONCLUSION AND DISCUSSION

This study aimed to use content-based filtering to recommend specific movies by considering genres and actors. We assessed the effectiveness of this approach by creating two content-based recommendation systems: a system based on the correlation between movies and genres and another based on the correlation between actors and genres.

We calculated the correlation among various genres based on actors and proposed a content-based recommendation system. The results of our study indicate that considering actors as a key component of a movie recommendation system helps suggest more appropriate movies to the users.

The findings of our study have several practical and academic implications for the South Korean film industry. From an academic perspective, the results of the study highlight potential future techniques that could be used when developing content-based recommendation systems. Moreover, because a number of hybrid recommendation systems include several procedures of content-based recommendation systems, the results of our study could also contribute to developing hybrid recommendation systems.

From a practical perspective, our study provides a notable foundation on which to build content-based recommendation systems for services with a variety of contents. Moreover, because only a few studies on user-customized contents have been conducted in the South Korean film industry, the results of this study can be employed to propose more concise and accurate recommendation services for potential customers. In addition, more customer-oriented perspectives could serve to further develop the South Korean film industry.

A. Limitation and Future Research

Although this study yielded several useful findings and implications, it has the following notable limitations. Because the dataset we used for our analysis only contained South Korean movies and actors, foreign films and actors who contributed to the South Korean film industry were not considered in the analysis.

Second, we only considered actor-based genre correlations for recommending movies. Because movies have a number of other important features, future research should include other features in the analysis.

Finally, we only employed South Korean movies that were released or being produced from 2010 to 2019, and thus, the results cannot be generalized. Moreover, considering that comparing other movie recommendation approaches was not fully examined in this study, future research would need to expand the comparison with other movie recommendation techniques.

REFERENCES

[1] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, Jun. 2005.

[2] P. B. Thorat, R. M. Goudar, and S. Barve, “Survey on collaborative filtering, content-based filtering and hybrid recommendation system,” *Int. J. Comput. Appl.*, vol. 110, no. 4, pp. 31–36, Jan. 2015.

[3] P. Resnick and H. R. Varian, “Recommender systems,” *Commun. ACM*, vol. 40, no. 3, pp. 56–58, 1997.

[4] R. Burke, “Hybrid recommender systems: Survey and experiments,” *User Model. User-Adapted Interact.*, vol. 12, no. 4, pp. 331–370, Nov. 2002.

[5] D. H. Park, H. K. Kim, I. Y. Choi, and J. K. Kim, “A literature review and classification of recommender systems research,” *Exp. Syst. Appl.*, vol. 39, no. 11, pp. 10059–10072, 2012.

[6] P. Melville, R. J. Mooney, and R. Nagarajan, “Content-boosted collaborative filtering for improved recommendations,” in *Proc. AAAI/IAAI*, vol. 23, Jul. 2002, pp. 187–192.

[7] G. Lekakos and P. Caravelas, “A hybrid approach for movie recommendation,” *Multimedia Tools Appl.*, vol. 36, nos. 1–2, pp. 55–70, Jan. 2008.

[8] C. C. Aggarwal, “An introduction to recommender systems,” in *Recommender System*. Cham, Switzerland: Springer, 2016, pp. 1–28.

[9] S. S. Li and E. Karahanna, “Online recommendation systems in a B2C E-commerce context: A review and future directions,” *J. Assoc. Inf. Syst.*, vol. 16, no. 2, p. 72, 2015.

[10] C. A. Gomez-Uribe and N. Hunt, “The Netflix recommender system: Algorithms, business value, and innovation,” *ACM Trans. Manage. Inf. Syst.*, vol. 6, no. 4, pp. 1–19, Jan. 2016.

[11] R. Zhou, S. Khemmarat, and L. Gao, “The impact of YouTube recommendation system on video views,” in *Proc. 10th ACM SIGCOMM Conf. Internet Meas.*, 2010, pp. 404–410.

[12] S. Reddy, S. Nalluri, S. Kunisetti, S. Ashok, and B. Venkatesh, “Content-based movie recommendation system using genre correlation,” in *Smart Intelligent Computing and Applications*. Singapore: Springer, 2019, pp. 391–397.

[13] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, “Item-based collaborative filtering recommendation algorithms,” in *Proc. 10th Int. Conf. World Wide Web (WWW)*, 2001, pp. 285–295.

[14] H.-N. Kim, A. El-Saddik, and G.-S. Jo, “Collaborative error-reflected models for cold-start recommender systems,” *Decis. Support Syst.*, vol. 51, no. 3, pp. 519–531, Jun. 2011.

[15] S.-M. Choi, S.-K. Ko, and Y.-S. Han, “A movie recommendation algorithm based on genre correlations,” *Exp. Syst. Appl.*, vol. 39, no. 9, pp. 8079–8085, 2012.

[16] T.-G. Hwang, C.-S. Park, J.-H. Hong, and S. K. Kim, “An algorithm for movie classification and recommendation using genre correlation,” *Multimedia Tools Appl.*, vol. 75, no. 20, pp. 12843–12858, 2016.

[17] E. Lawrence, “Everything is a recommendation Netflix, altgenres and the construction of taste,” *Knowl. Org.*, vol. 42, no. 5, pp. 358–364, 2015.

[18] M. Balabanović and Y. Shoham, “Fab: Content-based, collaborative recommendation,” *Commun. ACM*, vol. 40, no. 3, pp. 66–72, 1997.

[19] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, “Collaborative filtering recommender systems,” in *The Adaptive Web*. Berlin, Germany: Springer, 2007, pp. 291–324.

[20] M. J. Pazzani, “A framework for collaborative, content-based and demographic filtering,” *Artif. Intell. Rev.*, vol. 13, nos. 5–6, pp. 393–408, Dec. 1999.

[21] Y. Deldjoo, M. Elahi, P. Cremonesi, F. Garzotto, P. Piazzolla, and M. Quadrana, “Content-based video recommendation system based on stylistic visual features,” *J. Data Semantics*, vol. 5, no. 2, pp. 99–113, 2016.

[22] R. J. Mooney and L. Roy, “Content-based book recommending using learning for text categorization,” in *Proc. 5th ACM Conf. Digit. Libraries*, 2000, pp. 195–204.

[23] C. Basu, H. Hirsh, and W. Cohen, “Recommendation as classification: Using social and content-based information in recommendation,” in *Proc. AAAI/IAAI*, 1998, pp. 714–720.

[24] M. de Gemmis, P. Lops, G. Semeraro, and P. Basile, “Integrating tags in a semantic content-based recommender,” in *Proc. ACM Conf. Rec. Syst.*, 2008, pp. 163–170.

[25] U. Shardanand and P. Maes, “Social information filtering: Algorithms for automating ‘word of mouth,”’ in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 1995, pp. 210–217.

[26] C. Musco, C. Musco, and C. E. Tsourakakis, “Minimizing polarization and disagreement in social networks,” in *Proc. World Wide Web Conf. (WWW)*, 2018, pp. 369–378.

[27] A. De Vany and W. D. Walls, “Uncertainty in the movie industry: Does star power reduce the terror of the box office?” *J. Cultural Econ.*, vol. 23, no. 4, pp. 285–318, 1999.

[28] H. L. Vogel, *Entertainment Industry Economics: A Guide for Financial Analysis*. Cambridge U.K.: Cambridge Univ. Press, 2014.

[29] S. Basuroy, S. Chatterjee, and S. A. Ravid, “How critical are critical reviews? The box office effects of film critics, star power, and budgets,” *J. Marketing*, vol. 67, no. 4, pp. 103–117, Oct. 2003.

[30] E. Park, Y. Jang, J. Kim, N. J. Jeong, K. Bae, and A. P. del Pobil, “Determinants of customer satisfaction with airline services: An analysis of customer feedback big data,” *J. Retailing Consum. Services*, vol. 51, pp. 186–190, Nov. 2019.

Syjung Hwang received the B.Sc. degree from Sungkyunkwan University, Seoul, South Korea, in 2020, where she is currently pursuing the master’s degree with the Department of Interaction Science and the Data eXperience Laboratory.

Her research interests include cultural data science and application of artificial intelligence in our society.

Eunil Park received the Ph.D. degree from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2016.

He is currently an Assistant Professor with the Department of Interaction Science and the Founding Director of the Data eXperience Laboratory, Sungkyunkwan University. As one of the interdisciplinary scientists in the field, his research results have been published in numerous international social science journals as well as at scientific journals. His research interests include data science and data-driven user experience.

Dr. Park was the inaugural recipient of the NRF-Elsevier Young Researcher Award in Korea (Interdisciplinary Studies).