Analysing Gender Bias in IMDB Films Based on Social Networks

Jiayuan Chen¹, Minglu Cui² *

¹ Department of English, School of Foreign Languages, Northwest University, Xi’an, Shaanxi, 710000, China
²Department of English, School of Foreign Languages, Northwest University, Xi’an, Shaanxi, 710000, China
*Corresponding author’s e-mail: mtsuei96@nwu.cn

Abstract. The film industry has a major impact on society. This paper focuses on the gender bias in films. First, the social network analysis (SNA) method is used to construct a movie social network from the online movie dataset IMDb. By analysing these social network data, natural language processing methods are used to analyse movie titles, movie subtitles and casts. Then, this paper constructs a film gender bias classifier. Because the online movie data set IMDb lacks large-scale manual annotation data, and the classifier using traditional deep learning is prone to over-fitting, so this paper proposes a movie gender based on dual learning. The bias feature extractor and classifier enable training of the deep learning model from a small amount of annotated data and classify the bias. Finally, we find that women's roles in the film have improved in all aspects, including the central role of female characters.

1. Introduction
Gender bias is a potential problem in today’s society that permeates our culture, learning, workplace, and even our language, often in invisible forms. The first step in changing gender biases is to find out where they exist, and that is where the emerging field of computational linguistics comes in.

The film industry has a significant impact on society, as a microcosm of society, affecting aspects of life such as self-esteem and career choice. It is of great significance to analyse the portrayal of men and women in films and to dig out the gender bias in films.

Most previous studies on gender bias fall into two categories: 1. Provide simple statistics from the data to emphasize the gender gap [1]. 2. Use a small amount of data for data mining [2][3]. Carnes et al. [4] and Terrell et al. [5] analysed the film information retrieval tasks mainly involved, as well as the various previous works involved in this field. Fast et al. [6] focused on gender bias in English fiction, especially in online fiction writing communities. The work involved analysing the behaviour of men and women and describing it in this online novel. In addition, the study showed that men were overrepresented and found that this was traditional in the online fiction data used for analysis, and gender bias was common across all genres. Godden et al. [7] analysed children’s books and Millar et al. [8] These studies are useful for revealing trends, but the derivation of these analyses has been done on very small data sets. Kay et al. [9] analysed gender bias in occupations. In the image search results of occupations, compared with male counterparts, women are usually assigned lower level roles.

In the field of gender bias in movies, Lv et al. [10] established a social network of character movies, StoryRoleNet, to analyse gender bias in movies.
This study used machine learning and data mining algorithms to study gender bias in movies. At present, deep learning technology has achieved great success in computer vision, natural language processing and other fields, but it relies on large-scale manual annotation training data, and lacks large-scale manual annotation data set in the field of gender bias in movies. This project creatively used dual learning to mine the gender bias data of movies, and solves the problem of inadequate manual annotation data set in the field of gender bias in movies. The signal used by dual learning exists naturally in artificial intelligence tasks, but it is seldom used by people. We call it the structural duality of artificial intelligence tasks. Structural duality means that the output of one AI task happens to be the input of another, and vice versa. After solving the data set problem, this paper proposes an entity matching method to construct a movie social network from IMDb data of movie titles and credits for gender bias mining. Finally, this paper proposes a novel classifier that can predict whether a film has passed the Bechdel test and evaluate the changes in gender bias in thousands of films over decades.

2. Related Work

2.1. Film Social Network
In recent years, the study of social network has been widely spread. Researchers have found that the analytical techniques of social networks can be used in many areas that do not have explicit data with a network structure. The film industry, for example, is one such area. The researchers used social network analysis to analyze movies, gaining new insights not only about specific movies, but also about the film and television industry. For example, using social networks can empirically analyze the social connections between movie characters. Maintaining the Integrity of the Specifications.

Weng et al. [11] proposed RoleNet, a method for converting movies into social networks. The RoleNet algorithm builds the network by connecting links between characters that appear in the same scene. RoleNet is the basis for scene detection and face recognition based on image processing to discover the appearance of characters. Weng et al. evaluated RoleNet experimentally on 10 films and 3 television shows. RoleNet is used to perform semantic analysis of movies, find communities, detect protagonists, and identify story segments. Park et al. [12] developed character-net, another method for converting movies into networks. Character-net creates a social network for conversations between characters, and uses script caption alignment to extract who is talking to in the scene. Park et al. evaluated their method in 13 films. Character-net is similar to RoleNet, which is used to detect the protagonist and cluster the community.

Tran and Jung [13] developed CoCharNet, which added weights to links in interactive networks, where weights are functions of the number of times two characters occur together. Tran and Jung used CoCharNet to assess the importance of the characters in the movie. They demonstrated that network centrality (e.g., tightness centrality, inter-centrality, and weighting) can be used to categorize secondary and primary characters in films. For example, they used a near central character test with an accuracy of 74.16 percent. Lv et al. [10] proposed an algorithm to improve the accuracy of creating movie social networks. They proposed StoryRoleNet, which combines video and subtitle analysis to build a more accurate movie social network. Subtitles are used to add other links to the video analysis that may be missing. Lv et al. are similar to RoleNet and character-net. Use movie social networks to bring communities together and identify key characters.

2.2. Gender Bias Assessment
Many studies have attempted to assess the gender gap between men and women across different fields [14]. Researchers have identified many manifestations of gender bias in our society. Lariviere et al. [15] found that scientific papers in which women held lead author positions were less cited. Wagner et al. [16] observed the unequal coverage of men and women on Wikipedia. The situation for women in the film industry is similar to that in other fields: female characters are underrepresented and poorly portrayed [17]. The Bechdel test, named after Alison Bechdel's 1985 comic Dykes to Watch Out For,
aims to address the under characterization of female characters in films and television. Here are the tests:

1. There are at least two female characters in the movie (there are also variations that require female characters with names);
2. There is at least one conversation between two female characters;
3. This conversation can't be about men. Currently, only 57% of films pass this test. Researchers also used the Bechdel test. In recent years, studies have used the test to assess gender bias in movies.

Garcia et al. [16] quantified the Bechdel test and applied it to social media. They added YouTube trailers, movie scripts and Twitter data to produce 704 trailers that could play 493 movies. Garcia et al. [16] used movie data to create a conversation social network. They set up a network of conversations among Twitter users discussing the trailer. The Bechdel test score was calculated using the dialog network. The study found that movie trailers with a male bias were more popular. Similarly, they found that Twitter conversations had a similar bias to movie conversations. Agarwal et al. [18] examined the differences between films that passed and failed the Bechdel test. They created a classifier for the Bechdel test, which was trained in 367 films and evaluated in 90. They found that films that failed the Bechdel test tended to have women play unimportant roles.

2.3. Dual Learning
As a new machine learning paradigm, dual learning was first proposed by He et al. [19] in machine translation tasks. Dual learning solves the problem of insufficient training data in the practical application of machine learning. When we lack sufficient manually labelled large-scale data for effective training, we need to find other signals to drive the training process. The signals that dual learning uses are naturally present in artificial intelligence tasks, but are rarely used. We call this signal a structural dual property in an artificial intelligence task. Dual learning is proposed based on the observation that many machine learning tasks have dual forms. For example, in machine translation task, English translation into Chinese task and Chinese translation into English task are dual tasks; in image comprehension task, image generation text and text generation image are dual tasks; in speech recognition field, speech recognition and speech synthesis are dual tasks.

This research takes the dual learning in machine translation as an example, as shown in figure 1. In the machine translation task, it is assumed that we only have monolingual data, which are English documents with unmarked information and Chinese documents with unmarked information, as well as two weak initial 6.m and Chinese-English model. The task of dual learning is to use the unmarked monolingual data to continuously learn, improve the translation ability of the two initial models, and finally get a very strong two 6.s model and Chinese-English model. To achieve this, learn to use an English sentence with no information, translate it into Chinese using the original translation model, and then translate it back into English using the original translation model in the opposite direction. By comparing the original English sentence with the translated English sentence, as well as the grammar and morphology of the translated result in the middle, we can get a series of feedback signals to update the original model and make it continuously improve. When we have a large amount of monolingual data, dual learning can continuously improve the performance of the translation model to a high level.

![Figure 1. Machine Translation based on Dual Learning](image)

The idea of dual learning is not only limited to artificial intelligence tasks with dual tasks, but also applicable to artificial intelligence models with dual models. Such as automatic encoder (Autoencoder,
AE) against emergent network (Generative Adversarial Networks, GANs). Both the encoder and decoder in AE have structure dual properties, and the generator and discriminator in GANs also have structure dual properties. However, these dual properties of structure are often ignored in current studies. Encoders and decoders in AE are often trained separately in recommendation systems, and their feedback signals are often not shared. This paper proposes the design of stacked automatic encoder as a dual closed loop, which uses the structure dual property between encoder and decoder to train encoder and decoder simultaneously, and uses feedback signal to update the training process of encoder and decoder, so as to increase its performance of extracting commodity hidden layer features.

3. Gender Bias Approaches Based on Film Social Network

3.1. Film Social Network Construction

The key to constructing the movie social network is to construct an algorithm that can generate the movie social network directly from the input movie. This research used movie names and movie characters' name data to construct movie social network.

First, a movie social network is constructed by entering a movie $M = (V, E)$, where $V$ is the set of network vertices and $E$ is the set of edges that connect the vertices. Where each vertex $v \in V$ represents a character in the movie. Each edge $e = (u, v, w) \in E$ represents the interaction between role $u$ and role $v$ and $w$ refers to the times of interactions. As shown in figure 2, a movie social network is constructed using movie subtitles, and figure 2. (a) character names are extracted from movie subtitles using a named entity recognition algorithm. Figure 2. (c) to match character entities to the credits, and figure 2. (b) to link roles and increase edge weights by 1.

The steps to construct a movie social network through the movie credits and movie subtitles are as follows: 1. Use the named entity recognition algorithm to detect and extract individual or organizational entities from the input movie subtitles and store the occurrence time of such entities in the movie. 2. The entities in the movie credits are matched to the list of characters in the credits. It is worth mentioning that it is not possible to map one-to-one between characters in the cast character list and those extracted from the subtitles. In the dark knight, for example, Bruce Wayne is referred to as "Bruce Wayne" three times, Bruce 16 times and Wayne 20 times. 3. In order to solve the problem that the characters in the cast and crew character list and the characters extracted from the subtitle are not mapped one by one. This paper proposes a character entity matching algorithm. As shown in algorithm 1, first, divide all the characters into first and last names and link them to the full names of the actors and characters. Then, link all characters that appear in the role to the subtitle.

**Algorithm 1** Character Entity Matching Algorithm

| Data: | Actor Name, Character Name, Threshold Value; |
| **Results:** | Role Matching; |
| 1. Name $\leftarrow$ PersonName.split(); |
| 2. foreach $N_i \in \text{Names}$ do |
| 3. if $\text{Role}[N_i].length = 1$ then |
| 4. return $\text{Role}[N_i]$; |
| 5. end |
| 6. return $\text{MaxWRatio(PersonName, role}[N_i],\text{Threshold})$ |
| 7. end |

In order to solve the matching problem of actors' roles and names, as in algorithm 1. First, divide all the characters into first and last names and link them to the full names of the actors and characters (line 2). Then, if there is only one character with a specific last or first name (a one-to-one match), you need to link all the characters that appear in the subtitle to that character (lines 3-5). Compare strings
by using WRatio and match part of the name to the full name. With WRatio, the highest matching character with a score above the "threshold" is finally selected (line 6).

![Movie Social Network Construction Based on Movie Subtitle](image)

3.2. Social Network Features

In order to study gender bias in films, this study calculated five types of features in the film social network: vertex features, network features, film features, gender representation features and actor features. By calculation, we analyse how these characteristics change over time.

In addition, these characteristics are used to construct a machine learning classifier.

First, the definition of these features is given:

**Vertex characteristics:** the neighbor nodes of a vertex in a movie social network are defined as. The vertex feature is the total weight feature: the total weight of all edges, representing the number of times the character “v” appears in the movie $\text{Total}_w(v) = \sum_{(u,v,w) \in E} w$. Compactness center feature: the reciprocal of the total distance to all nodes in the graph, where, $C_v(v) = \frac{1}{\sum_{u \in V} d(v,u)}$ represents the shortest distance between them.

**Network features** include network edge number, network vertex number, network node group, and statistical network feature number.

**Movie characteristics** include: release time, movie IMDb ranking, total movie playing time, movie genre, movie IMDb rating.

**Gender represents characteristics:** number of genders, number of females in the first 10 roles. Number of actors, number of actresses, gender relations triple.

**Actor characteristics** include the year the actor was born, the year the actor died, and the age of the actor at the time the film was released.

In order to check the status of gender gap, the gender gap in movies, especially by type, was investigated. This study only analyzed the most popular movies (movies with more than n votes on IMDb). To answer the first research question, is there a genre without gender differences? First, the vertices and character features of all the characters are calculated, and then the data is divided by gender and movie type. Finally, Mann-Whitney U was used to test these characteristics to see if there were statistical differences between different types of male and female characters. To study
relationships in movies and answer the second question about what gender relationships reveal, this study calculated all relational triangles in the network and grouped them by the number of women in each triangle. Then, the triangle is subdivided by type and the change of triangle with time.

In order to investigate the role of gender centrality, the third research question of the centrality of female characters, this study calculated the PageRank of nodes in all movie networks, analyzed the number of men and women in the top 10 roles in movies, and studied how this number has changed over the years.

3.3. Gender Bias Feature Extractor

In order to achieve the film from a small amount of film subtitles with the movie ratings gender bias feature extraction. Using dual learning technology, this paper puts forward a kind of automatic encoder movie gender bias feature extraction based on duality, as shown in figure 3.

This paper proposes to use stacked automatic encoders to model movie subtitle data to extract hidden layer feature information from movie subtitle data. Traditional stackable automatic encoders do not consider the structural dual information between encoder and decoder, train encoder and decoder separately, and feedback signals between encoder and decoder are not shared. This paper presents a new type of automatic encoder structure, dual stack of first design and dual main encoder decoder structure is a kind of dual closed loop.

Then, the main encoder and the dual decoder are trained together, so that the feedback signals of the master encoder and the dual decoder can be shared, and the performance of the dual stack automatic encoder to extract the hidden layer feature information in the implicit data is improved.

As shown in the main encoder module in figure 3, the binary movie rating $e_i \in \mathbb{R}^n$ is first encoded by the main encoder into the hidden layer representation of the movie rating $e_i^r$. Where the main encoder code is expressed as:

$$
\text{Main Encoder}: \begin{cases}
e_1^i = a_1(W_1 e_i + b_1) \\
e_2^r = a_2(W_2 e_1^i + b_2)
\end{cases}
$$

(1)

Dual decoder decoding is expressed as:

$$
\text{Dual Decoder}: \begin{cases}
e_3^r = a_3(W_3 e_2^r + b_3) \\
n\hat{i} = a_4(W_4 e_3^r + b_4)
\end{cases}
$$

(2)

Among them, This subscript $i$ of $e_i^r$ represents the specific movie rating information, and the superscript $r$ represents the implicit representation encoded from the movie rating. $W_1 \in \mathbb{R}^{v_1 \times m}$, $W_2 \in \mathbb{R}^{v_1 \times v_1}$, $W_3 \in \mathbb{R}^{v_2 \times v_1}$, $W_4 \in \mathbb{R}^{v_2 \times v_2}$ represent the weight matrix. $m$ represents the number of comments, $v_1$ represents the dimension of the first hidden layer, and $v$ represents the dimension of the bottleneck layer.

The model modeling in this paper is based on the data of movie subtitle and movie rating. In order to better model the data information of movie subtitle and movie rating, this paper inserts confidence matrix into the square loss letter, and designs the model loss function of the dual automatic encoder of the gate attention mechanism as follows:

$$
L_{\text{GADAE}} = \sum_{i=1}^{n} \sum_{j=1}^{m} \|C_{ij}(D_{ij} - \hat{D}_{ij})\|^2 + \|C^T \odot (D^T - \hat{D}^T)\|^2
$$

(3)

Where $\odot$ represents the element product, $\| \cdot \|$ is the F-norm of the matrix, and the confidence matrix is defined as:

$$
C_{ij} = \begin{cases}
\rho & \text{IF } D_{ij} = 1 \\
1 & \text{ELSE}
\end{cases}
$$

(4)
In this paper, the objective function of the dual automatic encoder for the design of the gate attention mechanism is

\[
C_{i,j} = \begin{cases} 
\rho & \text{IF } D_{i,j} = 1 \\
1 & \text{ELSE}
\end{cases}
\]

\[
L = L_{\text{GADAE}} + \lambda \left( \|W_1\|^2 + \|W_2\|^2 \right)
\]

(5)

Figure 3. Movie Gender Gap Feature Extraction Based on Dual Autoencoder

**Algorithm 2** Dual automatic encoder training process

**Input:** Movie rating vector \(u_i\), Movie embedding vector \(e^c_i\), Fusion vector \(e^j_i\), Adjacent to the vector \(e^n_i\);

**Output:** Parameterized \(\theta_E\) main encoder \(E\), Parameterized \(\theta_D\) dual coder \(D\);

Random initialization \(\theta_E\) and \(\theta_D\)

repeat

- Get a minibatch of \(m\) pairs \(\{e_i, d_i\}_{i=1}^m\);
- Computing the gradient:
  \[
  G_E = \nabla_{\theta_E} \left( \frac{1}{m} \sum_{i=1}^{m} [L_E(f(e_i; \theta_E), y_i)] \right);
  
  G_D = \nabla_{\theta_D} \left( \frac{1}{m} \sum_{i=1}^{m} [L_E(f(d_i; \theta_D), y_i)] \right);
  
- update \(\theta_E\) and \(\theta_D\)
  \[
  \theta_E \leftarrow \text{Opt}_x(\theta_E, G_E), \theta_D \leftarrow \text{Opt}_x(\theta_D, G_D)
  \]
- until the model of convergence

The optimizer uses the master encoder and the dual decoder in the dual noise reduction automatic encoder adjacent to the attentional mechanism, so that the feedback signals of the master encoder and the dual decoder can be shared between the two models.
3.4. Film Gender Bias Classifier

To explore differences in the representation of men and women in the film industry, the study trained a classifier that used character data (vertex and actor characteristics) and movie data (IMDb characteristics without male and female counts) to predict a character's gender. The basic assumption is that if the representation of men and women is similar, the roles of men and women cannot be distinguished. But once you can predict the gender of a role, you can learn which characteristics are most descriptive.

To establish the data set, this study collected data on the most popular actors from the movies. And define a popular actor as an actor who has participated in at least N films and has more votes than m in IMDb. The idea is to avoid analyzing niche films and unknown actors with secondary roles and no meaningful function. For each actor's movie appearance, the above functionality was extracted in this study (see section 3.4.1) to evaluate the model due to the availability of movie participation data for each actor from multiple movies. This requires splitting all data for the same actor into training sets or validation sets. In order to solve this evaluation problem, this study used the method of setting aside a set of cross-validation. In order to create a true fact for the classifier, the gender of each actor must be determined. For most roles, extract the gender from the IMDb in a similar way. IMDb has actor or actress attributes, which allow us to identify the gender. As mentioned earlier, the IMDb dataset is only partial, so to overcome this problem, we use a dataset that maps the first name to the gender.

4. Experiment Settings

4.1. Data Selection

To evaluate and test our movie social network construction algorithm above on real data, we assembled a large dataset of movie subtitles and movie character lists. In addition, we collected a list of movie characters from the IMDb website and movie subtitles from 15,540 movies. In addition, we used data from 4,658 Bechdel films.

To analyze the content of the movie, we extract information from the subtitles. Subtitles are freely available online at many sites. For example, OpenSubtitles alone hosts more than 500,000 English subtitles that are manually created by the community. We collect subtitles using Subliminal 7, a Python library for searching for and downloading subtitles. Subliminal downloads the subtitles from multiple sources and uses internal scoring methods to determine which ones are best for a particular movie. Using Subliminal, we downloaded the subtitles for 15,540 movies.

4.2. Experimental Environment and Settings

The software and hardware environment of the experimental program in this paper is shown in table 1:

| Table 1. The Experimental Environment |
|---------------------------------------|
| Items                  | Details                      |
| Processor              | Intel(R) Core i9-9900k       |
| Graphics Card          | Nvidia GeForce GTX Titan X  |
| Development Language   | Python 3.6                   |
| Development Framework  | Pytorch                      |

5. Experimental Results and Analysis

To analyse the gender gap in the film industry, we analysed movie titles that received at least 1,000 votes on IMDb.

This results in a dataset containing 15,540 movies, which is 20 times the largest movie dataset currently available.

First, we predict the gender of the actors from the movie and the vertex features. We evaluated our categorized actors using cross-validation and found that we could predict gender by character
significantly better than by chance. The AUC of our classifier is 0.85, and the accuracy and recall rate are 0.76 and 0.79, respectively. Table 2 shows the ranking of the five most important features of the movie gender classifier.

Table 2. The Film Feature Index

| Features                  | Importance | Male Average | Female Average |
|---------------------------|------------|--------------|----------------|
| Actor age                 | 0.20280    | 44           | 45             |
| Actor's birth year        | 0.19913    | 1950         | 1960           |
| Film ranking              | 0.18571    | 6.43         | 6.4            |
| Release time              | 0.03254    | 1993         | 1995           |
| Broadcast time            | 0.02014    | 109          | 108            |

Then, in order to examine the relationships between the characters in the movie social network, we analysed the triangles in the relationship network. We found that most triangles have three people, we found the relationships between the roles, we analysed the role relationship triangles in the network. We find that most triangles have three men, while the triangle with three women is the least common, as shown in table 3, the proportion of women in the film social network triangle in the five film genres.

Table 3. Gender Classifier Top-5 Features

| The number of women in the triangle | 0   | 1   | 2   | 3   |
|------------------------------------|-----|-----|-----|-----|
| Action                             | 45.85% | 41.07% | 12.89% | 1.45% |
| Crime                              | 42.56% | 40.45% | 15.78% | 2.10% |
| Romance                            | 21.29% | 43.56% | 22.54% | 6.04% |
| War                                | 65.64% | 24.67% | 25.74% | 3.63% |
| Comedy                             | 56.98% | 41.98% | 20.84% | 3.97% |

In the five genres, the most common triangle type is 3 men (no women), and in all the others, 2 men and 1 woman. According to the results, romance was the most interactive type among women, while war was the least.

Next, we analysed the gender composition of the top 10 core characters, as shown in figure 4. We found that in most movies, more men than women play male characters. Moreover, as the data show, there are hardly any films without men and women in the top 10. Also, in many movies, the majority of the top 10 most important characters are men.

![Figure 4](image)

Figure 4. The distribution of movies by gender of the top-10 most central characters

6. Conclusion

In this paper, we proposed a new method of constructing movie social network, an algorithm of constructing movie social network based on character entity recognition, and a feature extraction algorithm of movie social network based on dual automatic encoder. Moreover, we constructed a
feature extraction algorithm of movie character gender bias and a feature classifier of movie character gender. Based on the IMDb dataset of online movie data, the experiment results show that gender differences still exist in almost all types of movies. For example, the relationship triangle in the movie is dominated by men. In terms of the top 10 leading movie characters, it’s mostly men. However, we also found that the majority of the top 10 most important roles were men. By the ecological analysis of IMDb films, we found that gender bias still exists. Films, as the major social media, female social gender stereotypes are always interpreted, which not only guides the audience, but also negatively affects the expression of female consciousness and the image and interpretation of women in films.

Acknowledgments
This research was part the project “Analysing Gender Bias in IMDb Films based on Dual Learning and Entity Matching”, which was supported by the sustentation item of innovation and entrepreneurship training program (S201910697354) for college students in Shaanxi province.

References
[1] Angela Coyle. Behind the Scenes: Women in Television[M]// Women and Work. 1988.
[2] Agarwal A, Zheng J, Kamath S, et al. Key female characters in film have more to talk about besides men: Automating the bechdel test[C]//Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2015: 830-840.
[3] Glascock J. Gender roles on prime-time network television: Demographics and behaviors[J]. Journal of Broadcasting & Electronic Media, 2001, 45(4): 656-669.
[4] Carnes M, Devine P G, Manwell L B, et al. Effect of an intervention to break the gender bias habit for faculty at one institution: a cluster randomized, controlled trial[J]. Academic medicine: journal of the Association of American Medical Colleges, 2015, 90(2): 221.
[5] Terrell J, Kofink A, Middleton J, et al. Gender differences and bias in open source: Pull request acceptance of women versus men[J]. PeerJ Computer Science, 2017, 3: e111.
[6] Fast E, Vachovsky T, Bernstein M S. Shirtless and dangerous: Quantifying linguistic signals of gender bias in an online fiction writing community[C]//Tenth International AAAI Conference on Web and Social Media. 2016.
[7] Gooden A M, Gooden M A. Gender representation in notable children's picture books: 1995–1999[J]. Sex roles, 2001, 45(1-2): 89-101.
[8] Millar B. Selective hearing: gender bias in the music preferences of young adults[J]. Psychology of music, 2008, 36(4): 429-445.
[9] Kay M, Matuszek C, Munson S A. Unequal representation and gender stereotypes in image search results for occupations[C]//Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM, 2015: 3819-3828.
[10] Lv J , Wu B , Zhou L , et al. StoryRoleNet: Social Network Construction of Role Relationship in Video[J]. IEEE Access, 2018:1-1.
[11] Weng C Y, Chu W T, Wu J L. Rolenet: Movie analysis from the perspective of social networks[J]. IEEE Transactions on Multimedia, 2009, 11(2): 256-271.
[12] Park S B, Oh K J, Jo G S. Social network analysis in a movie using character-net[J]. Multimedia Tools and Applications, 2012, 59(2): 601-627.
[13] Tran Q D, Jung J E. CoCharNet: Extracting Social Networks using Character Co-occurrence in Movies[J]. J. UCS, 2015, 21(6): 796-815.
[14] Jia S, Lansdall-Welfare T, Sudhahar S, et al. Women are seen more than heard in online newspapers[J]. PloS one, 2016, 11(2): e0148434.
[15] Lariviè re V, Ni C, Gingras Y, et al. Bibliometrics: Global gender disparities in science[J]. Nature News, 2013, 504(7479): 211.
[16] Wagner C, Garcia D, Jadidi M, et al. It's a man's Wikipedia? Assessing gender inequality in an online encyclopedia[C]//Ninth international AAAI conference on web and social media. 2015.

[17] Wood J T. Gendered media: The influence of media on views of gender[J]. Gendered lives: Communication, gender, and culture, 1994, 9: 231-244.

[18] Agarwal A, Zheng J, Kamath S, et al. Key female characters in film have more to talk about besides men: Automating the bechdel test[C]//Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2015: 830-840.

[19] He D, Xia Y, Qin T, et al. Dual learning for machine translation[C]//Advances in Neural Information Processing Systems. 2016: 820-828.