Research Article

Analysis of Sports Popular Trend Based on Public Opinion Mining of New Media

Qingtuan Xie, Qing Han, and Dong Chen

1 China Wushu School, Beijing Sport University, Beijing 100084, China
2 College of Physical Education and Sports, Beijing Normal University, Beijing 100875, China
3 China Athletics College, Beijing Sport University, Beijing 100084, China

Correspondence should be addressed to Dong Chen; 2019980001@bsu.edu.cn

Received 14 May 2022; Revised 29 June 2022; Accepted 1 July 2022; Published 17 August 2022

1. Introduction

In recent years, there have been numerous studies on public opinion, most of which are based on the perspectives of communication, journalism, and public opinion [1]. However, there are very few articles combining the sports perspective. With the rapid development of my country’s media and the rising level of national sports competition, citizens and media reporters are increasingly paying attention to sports [2]. At present, China’s online social networks have a wide audience, and the number is still increasing. With the further expansion of the popularity of the Internet, the public opinion influence of China’s online social networks began to increase significantly [3]. The realization process of topic identification and tracking includes five subtasks, namely, report segmentation, new event identification, relationship detection between different reports, topic identification, and topic tracking. Among these five subtasks, the topic identification and topic tracking subtasks are highly valued, and there are many related researches [4]. The superposition of a variety of factors has caused the public opinion to present diversified characteristics. In particular, new media such as the mobile Internet have gradually intervened in the daily life of the public, changing the original public opinion ecology, and making the development of various public opinions show different characteristics and laws. These different public opinions not only may promote the escalation of mass events but also may turn various crisis events into opportunities due to proper guidance [5]. Sports are inseparable from communication, and the communication industry is also inseparable from the sports industry. The more developed the society, the closer the relationship between sports and communication.
With the continuous development of communication technology, the advantages of new media, such as fast communication speed, wide communication range, and information dissemination, have increasingly penetrated into all fields of social organizations and social life, prompting sports to develop from the original amateur behavior to the professional and commercial direction [6]. As an important part of network public opinion, the current network public opinion of sports events also faces the problems of “group imbalance and complex demands of expressors, platform fragmentation and dissemination bias, and differentiation of opinions and lack of consensus” [7]. Therefore, understanding the essence and performance characteristics of online public opinion of sports events in the new media environment has become the premise and foundation of building scientific, reasonable, and effective governance rules of online public opinion of sports. Although there are reasons for the emergence of these public opinion crises due to the management system and the personal quality of relevant sports practitioners, from the perspective of public opinion intervention, the role of the rapidly developing online media in fueling the topic is also an important reason for the rapid spread of the topic and attracting attention [8]. Sports communication is moving towards a brand-new communication pattern with audience as the core, technology as the driving force, and Internet as the platform [9]. In addition, since it is impossible for online news users to browse all news reports, people only pay attention to what they are interested in or just hope to get the news topics that are currently attracting more attention. That is to say, not all news topics will attract people’s attention or get widely concerned and discussed [10]. The Internet has driven the popularization of sports information and facilitated the attention, forwarding and comment of netizens on hot spots and major sports events. The openness of the network platform also provides a channel for the generation, accumulation, and diffusion of public opinion on sports events.

Therefore, under the profound influence of sports network public opinion, the social effect caused by sports crisis network public opinion communication will also produce “ripple reaction.” The spillover effect of sports crisis network public opinion communication will not only affect the sports industry and other related industries but also spread to a larger space [11]. New media is an important factor in the development of online public opinion events. After collecting and analyzing sports network public opinion events, this paper finds that part of the role of the media has been strengthened and highlighted in network public opinion events [12]. Different from the presentation in news events, according to the different development stages of network public opinion events, the social situation of the media has changed, and individual media roles have been highlighted correspondingly with a set of specific behavior patterns.

2. Methodology

2.1. Connotation and Essence of Sports Public Opinion. Public opinion refers to the public opinion on the situation of the people in a relative social space around how social emergencies arise, change, and solve, as well as people’s personal attitudes and views on government personnel and managers of related events [13]. This is also the sum of the attitudes and views expressed by the broad masses of the people towards different events and problems in society. In online public opinion events, the development of news practice activities is carried out by the mass media as a socialized organization, through the formulation of industry standards, rational allocation of personnel division of labor, organization and coordination of work in various links and other role behaviors, as well as socialized executors. It is jointly realized by the individual subjective initiative of news practitioners [14]. The research shows that the public opinion detection technology based on pattern recognition cannot well adapt to the information law of different information sources, so it has great limitations and can only carry out small-scale fixed-point monitoring. With the continuous rise of our media platform and China’s continuous undertaking of international sports events, the phenomenon of sports public opinion crisis communication in China is also becoming more and more frequent [15]. The whole process of sports public opinion crisis communication is a process of information transmission and sharing, and a two-way circulation process of information. Generally, there are three kinds of public opinion in sports events: the first is implicit public opinion, the second is explicit public opinion, and the third is to express public opinion [16]. The following Figure 1 is a general sports information search mode diagram.

At the same time, the focus of public opinion on sports events is not focused, and it is easy to change the focus with the change of events. This requires us to effectively guide the public opinion of sports events [17]. First, we need to create a good public opinion atmosphere for sports events; second, we should make a real and effective analysis of the whole event report of public opinion in sports events so as to prevent the chaos of public opinion to a certain extent; finally, we should take the correct three outlooks as a guide to make sports public opinion develop in the right direction. The official media, opinion leaders, public figures, network celebrities, and so on have innate advantages in mastering sports discourse resources, and their influence, control, and guidance on netizens’ public opinion are constantly strengthened, which leads to differentiation. Nan Zhenhong, the former editor-in-chief of Xinhua News Agency, once put forward the concept of “two public opinion fields,” which was used to distinguish the oral public opinion from the official public opinion.

In the context of new media, the public opinion of sports events has the functions of dissemination, guidance, communication, and supervision [18]. The communication function of public opinion in sports events is simply a kind of information, that is, what happens, what consequences will there be, and people’s views [19]. The media has complex and diverse role functions in the network public opinion events. These functions and the role expectations and behaviors attached to the media in the role-playing process discussed in Section 3 run through the whole process of the development and changes of the network public opinion events.
2.2. Identification and Analysis of Public Opinion in Sports Events. The construction of topic identification and tracking system mainly includes three parts: topic modeling, similarity calculation, and threshold estimation. Among them, modeling is the main factor affecting performance [20]. There are core reports in an event and core events in a topic, and the importance of core reports in an event should be higher than other reports, and the importance of core events in a topic should be higher than other events. Compared with traditional media, modern online media, digital media, touch media, mobile media, and other new media have won the attention of modern audiences because of their strong communication advantages and distinctive communication characteristics. At present, the existing classical topic models include the vector space model and probability model, among which the widely used probability models are language model and Dirichlet distribution [21]. The vector space model in the field of information retrieval is applied to topic modeling. The basic idea is to use feature vectors to represent news reports and topics. These features are mainly verbs, nouns, adjectives, and numbers in news reports. Although place names are represented by multiple words, they are treated as a feature [22]. In this article, the sports report is set to \( S \) and its vector expression is as follows:

\[
S = (k_{s,1}, w_{s,1}; k_{s,2}, w_{s,2}; \ldots; k_{s,n}, w_{s,n}),
\]

where \( k_{s,i} \) represents the features selected in the sports report \( S \) and the weight corresponding to \( k_{s,i} \) is \( w_{s,i} \). At this time, the weight needs to be calculated as follows:

\[
w_{s,i} = \frac{tf_i}{df_i + 2} \times (1 - \log \sqrt{df_i}),
\]

where \( w_{s,i} \) represents the weight of feature \( i \) in new sports reports, \( tf_i \) represents the number of times feature \( k_i \) appears in new, \( df_i \) represents the number of reports in which feature \( i \) appears, and \( N \) represents the total number of reports, so the vector of \( t \) at this time is expressed as

\[
t = (k_{t,1}, w_{t,1}; k_{t,2}, w_{t,2}; \ldots; k_{t,n}, w_{t,n}).
\]

At this time, the inner product of the report vector and the topic vector should be calculated, and the similarity between the two is

\[
sim(s, t) = \frac{\sum_{i=1}^{N} w_{s,i} \times w_{t,i}}{\sqrt{\sum_{i=1}^{N} w_{s,i}^2 \times \sum_{i=1}^{N} w_{t,i}^2}}.
\]

Given the threshold \( \theta > 0 \), if \( \sim(s, t) \geq \theta \), the report \( s \) is considered to belong to the topic \( t \) at this time; otherwise, it is considered that the report \( s \) does not belong to the topic. Probabilistic topic model is based on LDA theory. Local density approximation (LDA) is one of the approximations used in the exchange dependent energy functional of density functional theory. Local density approximation (LDA) is one of the approximations used in the exchange dependent energy functional of density functional theory. This approximation considers that the exchange correlation energy functional is only related to the values of electron density at various points in space. Language model is a probability-based model. The N-yuan language model is widely used in the field of information retrieval and information filtering, but in the field of topic recognition and tracking, the unary language model is generally adopted. Figure 2 shows the basic model framework of LDA.

The LDA model assumes that the topics are independent of each other and cannot obtain the relationship between the topics, which is not in line with the actual situation. Then, sample topologies of the LDA model, the CTM model, the four-layer PAM model, and the arbitrary PAM model are proposed. The main PAM model diagram is shown in Figure 3.

Each leaf node is a word, and nonleaf internal nodes are a topic. Each topic is based on the Dirichlet distribution of its leaf nodes. Each Dirichlet samples a polynomial, starts from the root node, samples its child nodes according to the polynomial distribution, and samples along the path of the directed acyclic graph until the leaf node produces a word. Let \( C_{\text{miss}} \), \( C_{\text{fa}} \), represent the cost of omission and false alarm, respectively, and \( P_{\text{target}} \) represent the probability of finding a new report. Table 1 evaluates the default values of cost parameters for different subtasks.

You can get an indicator to identify the cost:

\[
C_{\text{det}} = C_{\text{miss}} \times P_{\text{miss}} \times P_{\text{target}} + C_{\text{fa}} \times P_{\text{fa}} \times P_{\text{non-target}}.
\]
At this time, it can be obtained that the calculation formula of the false alarm rate and the false alarm rate is

\[ P_{\text{miss}} = \frac{c}{a + c}, \]
\[ P_{\text{fa}} = \frac{b}{b + d}. \]  

The meanings of four parameters are shown in Table 2.

### Table 2: Meaning of the four parameters.

<table>
<thead>
<tr>
<th></th>
<th>&quot;Yes&quot; is correct</th>
<th>&quot;No&quot; is correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>The system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>determines &quot;yes&quot;</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>The system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>determines &quot;no&quot;</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

the similarity between the document and the query. The public opinion of sports events is generally divided into the third stage [23]. The development of things follows the law of cause and effect, not out of nothing. It often has a certain “foundation,” “condition,” or “background” before the “start.” This stage is the “preparation stage” before the outbreak of network public opinion events—the incubation period of public opinion. In the absence of a more moderate transition period, the amount of information accumulates and rises rapidly. Therefore, after the incubation period, the network public opinion often ushered in a rapid outbreak period and pushed the trend of public opinion to meet the first peak. From the turbulent state of the public opinion field during the outbreak of public opinion to the obvious differentiation of opinions during the climax of public opinion, the public opinion field at this time has been basically completed. It is the result of the continuous rise, accumulation, and dispersion of public opinion after the outbreak, and it is the situation in which the collision of public opinion reaches the most intense state. Figure 4 shows the basic inference network model.

Although in the era of current media, information communication has shown obvious characteristics of "decentralization," the "decentralization" mentioned here is not without centralization, but the dilution or weakening of the center. Its remarkable performance is that in the information dissemination of some sports tourism crisis events, netizens become the main publishers of news or information, and some news or information provided by netizens become the source of traditional media and news websites. Because the current media channels are diversified, popular, grassroots, and truly comprehensive, eventually all the content of sports public opinion will be instantly presented in the public's field of vision regardless of the subjective wishes of the parties involved in the event, and this trend will continue be strengthened. After the outbreak of public opinion, public opinion continues to accumulate, and there are public opinion sets with different opinion orientations, including the same or similar opinion sets and completely opposite sets. When there is controversy, there will be discussion. These public opinion collections collide, melt, and converge with each other to jointly promote public opinion to reach its peak and reach a relatively saturated state. The evolution from the old media to the new media is bound to go through the development trend of humanization; that is, the degree of media technology continuation depends on how much it copies the natural model. The development of media is the result of human’s natural selection. At present, Weibo presents a serious situation of "fragmentation" and “differentiation,” which is the communication effect brought by the new communication environment, and it can achieve accurate communication and accurate measurement. Therefore, clients and operators have the potential advantage.
of touching audiences and markets more. Among them, there are three main ones: mainstream media leads the main line of sports public opinion crisis communication, Weibo creates a hotbed for sports public opinion crisis communication, and interpersonal communication sports public opinion crisis communication.

2.4. Analysis and Research on Mining Public Opinion Information Based on Sports Events. This paper uses a variety of different algorithms to analyze the popular trend of sports. The centroid algorithm locates unknown nodes based on network connectivity. It is very simple and easy to implement the localization algorithm without the cooperation of anchor nodes and unknown nodes. For those applications that do not require high positioning accuracy, the centroid algorithm is a good positioning method. The centroid algorithm is to uniformly describe documents with different lengths. Each document is normalized to a unified length and set to \( \| S \|_2 = 1 \). Let \( C = \left\{ C_i \right\}_{i=1}^l \) represent the predefined category set, and the centroid of category \( C_i \) is defined as

\[
C_i = \frac{1}{\left\| C_i \right\|_2} \sum_{S \in C_i} S.
\]

The similarity between the vector and the centroid vector at this time can be expressed by the cosine function:

\[
\cos(\vec{S}, \vec{C}_i) = \frac{\vec{d} \cdot \vec{C}_i}{\| \vec{S} \|_2 \times \left\| \vec{C}_i \right\|_2} = \frac{\vec{d} \cdot \vec{C}_i}{\left\| \vec{C}_i \right\|_2}.
\]

To judge the type of \( d \), we can calculate the normalized vector \( \vec{d} \) of \( d \), then calculate the similarity of \( \vec{d} \) and \( L \) centroids, respectively, and finally classify \( d \) into the category with the greatest similarity, that is, the category \( d \) of \( c \) is given as

\[
c = \arg \max_{c_j} \left( \cos \left( \vec{d}, \vec{C}_j \right) \right).
\]

The term base of the report is constructed by using the sample report. The general process can use the word segmentation technology. If there is a repeated term, the weight of the term is represented by the average of the weight. The expression is generally as follows:

\[
w(s,k_j) = \frac{\text{freq}(s,k_j) + 0.5N_{\text{begin}} + 0.5N_{\text{end}} + N_{\text{title}}}{\sum \text{freq}(s,k_j)},
\]

where \( w(s,k_j) \) is the weight of the term \( s,k_j \); \( \text{freq}(s,k_j) \) is the frequency of occurrence of the term \( s,k_j \) in the report \( s \); \( N_{\text{begin}}, N_{\text{end}}, N_{\text{title}} \) represent the frequency of the term in the beginning, end, and title of the report \( s \), respectively; and \( \sum \text{freq}(s,k_j) \) means the sum of the frequencies of all terms in \( s \). When the term set is further simplified, the event term set \( E \) can be obtained, and the set \( m \) composed of the first \( S' \) feature items with larger weight is reselected from the selected feature options for reporting to describe the event \( e_x \), which is expressed as

\[
e_x = \cup S',
E = \cup e_x = \{ek_1, ek_2, ..., ek_n\}.
\]

The weight of the term \( e_x \) in the collection \( s, k_q \) is mainly calculated according to the frequency of its occurrence in the reports, and all reports are carried out around the core or mainstream reports. Therefore, it is necessary to improve the weight of the terms in the reports. Therefore, there are

\[
w(e_x,k_q) = \alpha \times w(s,p,k_q),
\]

where \( w(e_x,k_q) \) is the weight of the term \( e_x \) in the event \( e_x,k_q \); \( w(s,p,k_q) \) is the weight value of the term \( s,p \) in the core report \( s,k_q \); and \( \alpha \) is an improvement parameter with a value greater than 1. For the term \( t \) describing the topic \( k_j \) comes from the core event, the weighted event is

\[
w_{ij} = \beta \times w(e_x,k_q).
\]

Any topic has its timeliness, and sports public opinion is no exception. Therefore, the timeliness analysis of the topic is also an important part of the general trend analysis. The heat of the news report will gradually decrease with the passage of time, and the heat of the topic of the news report will also decrease. This is the natural process of the gradual decline of the topic heat, that is, the news topic has timeliness. Table 3 shows the characteristic distribution.

In order to express the recession process, this paper uses the recession surplus ratio \( \text{rest}(t) \) to express it, and its concrete expression is

\[
\text{rest}(t) = e^{-\lambda t} (t \geq 0).
\]

Among them, the time when the topic’s popularity decline is denoted by \( t \), and the speed of the topic’s popularity decline is denoted by \( \lambda \). The remaining ratio \( t \) of the decline of topic popularity will decrease as the value of XXX increases. Therefore, the decay rate of topic heat also depends on the user’s interest. In this paper, the definition range of interested user is

\[
I(\text{topic}) = \frac{N(\text{comment/topic})}{N(\text{comment})}.
\]

The number of user comments included in the definition scope is expressed by \( N(\text{comment/topic}) \), and the sum of comments on all topics is expressed by \( N(\text{comment}) \). When
more users are interested in a topic, the popularity of the topic will decline at a slower rate.

In this part, this paper expounds the selection of news text feature items and the weight calculation of feature items, making full use of important information in the title, and increasing the weight of named entity information, so as to build a more accurate representation model of news text. Second, this section introduces the named entity vector into the calculation of text similarity and improves the method of calculating text similarity. The first factor affecting the performance of model recognition is the representation of new reports. The more the terms in the new report vector, the more the query information. If all these information are useful, the recognition performance can be improved. If there are too many terms in the report vector, there may be redundant information, which will reduce the recognition performance.

3. Result Analysis and Discussion

In order to improve the trend analysis of sports events based on new media public opinion mining, this paper will conduct experimental analysis on the conceptual model designed above. Establishing a scientific, accurate, and feasible trend analysis model is the premise and foundation of accurate experiments. By observing the data, it is confirmed that the model can well match the analysis of sports popular trend. Therefore, this paper will analyze it from two aspects: DCG value comparison and trend tendency accuracy analysis. Figure 5 shows the comparative analysis of DCG values on top-3, top-5, and top-10.

As can be seen from Figure 5, the average DCG value of the extended model is higher than that of the basic model. Generally, few query words are input by users; that is, each user uses an average of 1.75 words to describe the query, which often fails to express the query intention accurately and completely. The basic model only considers the matching degree between the query term and the indexed terms of each document, which may result in the omission of related documents that have a low matching degree with the query term but are higher than the actual query demand of users. Before and after the experiment, the accuracy of trend tendency was also analyzed, which is an important factor to grasp the main body of sports analysis. The data graph after the experiment is shown in Figure 6.

The accuracy of the trend analysis of the model before and after the experiment. According to the experimental results, the data diagram is drawn, as shown in Figure 6, where A represents the performance of the model without trend analysis experiment, and the implementation part represents the performance of trend analysis after experiment A. Under the same threshold, the closer the points in the graph are to the horizontal axis, the lower the accuracy of the trend analysis, and the better the performance. Observing Figure 6, it can be found that the performance curve after using the experiment A is closer to the origin of the coordinates, so its performance is better than that before using the model experiment. The experiment is aimed at the disadvantage that the trend of sports public opinion analysis and tracking is not clear. The optimization method is proposed, and the accuracy is improved by 87.5%. It is not only applicable to the trend analysis model proposed in this paper but also can be used to optimize other dynamic topic models.

4. Conclusions

The evolution of new media technology, the value appeal of the media platform and the audience’s differentiation on the production and consumption of sports information together constitute the internal tension of sports event network public opinion, and the guidance of sports network public
opinion must be based on this logic. Traditional media and new media can interact together to realize the interaction in the form and content of sports information dissemination. The complementary development of sports TV media and new media will bring about comprehensive changes to my country’s sports TV channels, which is the trend of my country’s sports media development. Focusing on the related research involved in information retrieval, this paper summarizes the belief network retrieval model, vector space model, and document relationship, analyzes their shortcomings, and then puts forward the research content of this paper. In the new media era, the communication mode of sports public opinion crisis is shown in three aspects: first, the communication mode is led by mainstream media; second is to use Weibo’s self-media platform to create a “hotbed” of sports public opinion communication; and third, the spread of sports public opinion crisis is realized through interpersonal communication through the network. In the experiment, the optimization method was proposed in view of the lack of clear tracking trend of sports public opinion analysis, and the accuracy was improved by 87.5%. However, the study still has some limitations. The research did not summarize the belief network retrieval model, vector space model, and document relationship. Therefore, there is still much room for development in public opinion analysis. The research needs further analysis.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

References

[21] F. Wunderlich and D Memmert, “Innovative approaches in sports science—lexicon-based sentiment analysis as a tool to
