ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202003_3(3).0022

Empirical Analysis on User Profile Model In University Library

Jiaojiao Li[1, a]

[1]China Academy of Information and Communications Technology, Beijing, 100000, China.

[a]ljiaoer@163.com

Abstract

As one of the tools to provide accurate services in the era of big data, user profile has attracted widespread attention. In order to improve the personalized service ability of university library, the construction of user portrait model is imperative. Based on the data of university library, this paper uses k-means algorithm to analyze the user profile. Finally, taking the relevant data of Fudan University Library as an example, this paper analyzes each cluster of user profile, and finally puts forward corresponding suggestions for the development of the library according to the results of user profile.

Keywords

Library; user profile; K-means.

1. Introduction

Because of the impact of Internet technology, the utilization of book in university libraries is facing severe challenges, and the utilization rate is on a decline . Taking Fudan University Library as an example, the number of books borrowed in 2013 reached 420000, while the total number of books borrowed in the last three years (2016, 2017, 2018) is 320000, 29000, 320000 in turn. In order to improve the library lending states, with the help of the user profile technology, this paper analyzes the general attributes (gender, age, education background, occupation, etc.) and behavior information (when, where, who and how) of the users in the university library, and then recommends books that are more suitable and meet the personalized needs of the users to further improve the borrowing volume of university libraries.

Whether from the perspective of the library or users, the implementation of library user profile is imperative. From the perspective of users, user profile can help the library to more targeted positioning of readers, and then provide them with books more suitable. From the perspective of the library, user profile can help the library to better understand the users' habits, preferences, needs, and improve service mode, enhance user experience, and improve the utilization rate of library resources. Therefore, the construction of user profile model is of great benefit to the library and users. This paper will take the borrowing data of Fudan University Library as an example for empirical study.

2. Related Work

Cooper, the father of interaction design, first put forward the concept of user profile, which is also called user role. It is mainly a user model constructed by means of questionnaire and interview. It is a virtual representative based on the real data of users [1]. Then the user profile is generated in the big data environment, which is mainly based on the user's social attributes (gender, age, education, occupation, residence, etc.) and behavior information (click, browse, collection, etc.), using science technology and data mining algorithm to abstract user needs or preferences and uses tags to represent the process, that is, to analyze the user attributes and user behaviors to characterize the same type of users from different dimensions [2,3].

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202003_3(3).0022

Now, Most scholars have been building the user profile model of digital library, rather than university library. Tejeda Lorente et al. Use the fuzzy language method and bibliometrics to build the user profile model and reduce the cold start problem in the recommendation system of digital library [4]; Jomsri uses the association rule algorithm to build the user profile model and further build the book recommendation system of digital library [5]; Kuzelewska builds the user profile model In this paper, a hybrid recommendation system is proposed, which uses clustering method to find the similarity between users, and puts forward the technology of identifying user profile[6]; Kovacevic uses clustering and classification prediction method to build user profile, and realizes the personalized recommendation of Digital Library [7]. In addition, a few of scholars build traditional library recommendation system based on user profile model. Shirude has developed an efficient library recommendation system based on multi-agent, which recommends the appropriate resources (literature, books) in the library to users through user interest profile [8]. Hu Yuan et al. Conducted modeling analysis on the user profile of the digital library [9]. Chen Tianyuan, starting from the user preference of university library, constructs the user profile model based on label by using factor analysis, cluster analysis and discriminant analysis [10]. Taking the National Library as an example, Yang Fan proposed to build a library big data analysis platform based on readers' and resources' profiles [11]. He juan uses the data of readers' borrowing behavior to construct the user profile model through the vector space model, and then uses the clustering algorithm to cluster the users to further realize the personalized recommendation of books in the University Library [12]. Based on the concept lattice, Xu hailing and others accurately portrayed the group user profile of University Library [13].

In a nutshell, the construction of the user profile model of university library has a certain research foundation. Based on the real data, this paper will make an empirical analysis of university library, and put forward corresponding suggestions for the accurate service of the library according to the results of the user profile.

3. Research Methodology and Data

3.1. Methodology

The K-Means algorithm was put forward by Macqueen. It was proposed in 1967, which belongs to unsupervised learning algorithm. The main principle of k-means is: for a given data set, the similarity can be judged by the distance between samples in the data set. Therefore, the data set is divided into K clusters. The distance within the group is as small as possible, and the distance between groups is as large as possible, that is,

there are large differences between samples in the same cluster, but not There is little difference between the samples of the same kind.

The step of K-Means algorithm is as follows:

Suppose that there are N samples $X = \{x_1, x_2, x_3, \dots, x_n\}$. To divide N samples into K different classes so that objects with high similarity can be divided into the same classes, the specific implementation process is as follows:

Step1: Select the cluster center. Randomly select K samples from a given set of samples as the initial clustering centers $c_1, c_2, c_3, \dots, c_n$;

Step2: Calculate the similarity. Based on the selected cluster center in the step1, calculate the distance between the rest of the sample and it. According to the principle of minimum distance $d_j = \|X - c_j\|$, assign a sample to each initial cluster center, where $X = \{x_1, x_2, x_3, \dots, x_n\}, j = 1,2,3, \dots, K$, complete the first iteration;

Step3: Select a new cluster center. According to the result of clustering in the step2, calculate the new clustering center by calculating the average value, that is, according to the formula:

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202003_3(3).0022

$$c_j = \frac{1}{n_j}\sum_{i=a}^{n_j} X_j$$

Where n_j is the number of samples of class j, $j = 1,2,3,\ldots,K$;

Step 4: Iterate continuously. Repeat steps 2 and 3 and iterate continuously.

Step 5: Calculate the sum of squared error and criterion function. Calculating the clustering criterion function when the clustering center is constant.

$$E = \sum_{j=1}^{k}\sum_{i=1}^{n_j}\left\|X_i^{(j)} - c_j\right\|^2$$

Among them, S_j is the cluster, c_j is the cluster center of S_j, $X_i^{(j)}$ is a sample of cluster S_j. When the clustering criterion function converges, the iteration stops, otherwise continue to repeat step 2 and step 3.

3.2. Data

This paper will select the borrowing data of Fudan University library from 2017 to 2018 academic year, and calculate the borrowing amount of each student in different periods (in the morning, in the afternoon, in the evening, at the beginning of the semester, in the semester, at the end of the semester), the collection land with the most borrowing times, the borrowing amount of the most sub-library, and the borrowing amount of nearly six months. Then, data processing is carried out to delete the records with more missing values and deal with the abnormal values of each data. After the above data sorting, merging and data cleaning, there are 14967 effective records, and 14967 students.

4. The Experimental Study of User Profile

4.1. The Construction of User Labels

Table 1. User Data

Type	Tag	Tag Abbreviation
Who	Reader Number	ReaderNo
	Reader Sex	ReaderSex
	Reader Department	ReaderDept
	Reader Status	ReaderStatus
	Reader Grade	ReaderGrade
What	Total books borrowed	TotBorNum
	Total books renewed	TotRenNum
	Total books Subscription	TotSubNum
When	Book borrowing in the Morning	MorBorNum
	Book borrowing in the Afternoon	AftBorNum
	Book borrowing in the Evening	EveBorNum
	Book borrowing at the beginning of term	TermStartBorNum
	Book borrowing at the Middle of term	TermMidBorNum
	Book borrowing at the End of term	TermEndtBorNum
Where	The Category of Sub-library	CateSubLib
	The Most Borrowed Sub-Library	MSubLibBorNum
	The Most Frequently Borrowed Sub-Library	MSubLib
Active State	The Borrowing of nearly six months.	SixMotBorNum

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202003_3(3).0022

Based on the above data, data import and calculation are carried out. Firstly, this paper will construct user labels, which conclude who, when, where, what, and active state. The main data structure is shown in Table 1.

1) Who, mainly including readers' number, reader's gender, reader's department, reader's education background and reader's grade. Among them, reader departments mainly include 58 departments, such as Department of materials science, big data, affiliated pediatric hospital, law school, etc., which are divided into three categories: Arts Department, Science and Engineering Department, and Medical Department; reader education is divided into undergraduate, master and doctor. Readers are mainly divided into grade one, grade two, grade three, grade four and above.

2) What, mainly includes readers' total borrowing amount, total renewal borrowing amount, and total subscribing amount. The total borrowed amount represents the total amount of books borrowed by the reader, the total renewed amount represents the total renewed amount of books renewed by the reader, and the total reserved amount represents the total amount of book reserved by the reader.

3) When, mainly including the amount borrowed in the morning, in the afternoon, in the evening, at the beginning of the semester, during the semester and at the end of the semester. The amount borrowed in the morning refers to the total amount borrowed by readers from 6:00 a.m. to 12:00 p.m. in a year, the amount borrowed in the afternoon refers to the total amount borrowed by readers from 13:00 p.m. to 18:00 p.m. in a year, the amount borrowed in the evening refers to the total amount borrowed by readers from 19:00 p.m. to 6:00 a.m. in a year; the amount borrowed at the beginning of a term mainly refers to the total amount borrowed by readers within two months of the beginning of a term. The borrowing amount during the semester mainly refers to the total borrowing amount of readers during the semester. The amount borrowed at the end of the semester refers to the total borrowing amount in the last two months of semester.

4) Where, mainly includes the sub-library with the most borrowing times, the classification of the sub-library, and the borrowing amount of the sub-library. The sub-library with the most borrowing times mainly refers to the readers borrowing books from different sub-library, and screening out the sub-library with the most borrowing amount, and the corresponding borrowing amount. At the same time, the sub-library are divided into Arts sub-library, Science sub-library, and Medical sub-library.

5) Active State. Active State is mainly reflected by the total amount of borrowing in the past six months. Total borrowing in the past six months refers to the total borrowing from March 2018 to August 2018.

4.2. Empirical Study of User Profile

Based on the user's character characteristics, events, time, place and active status, this paper uses K-means algorithm to make profile of the user group.

1. Data transformation

Based on the borrowing data of Fudan University Library, discrete data is converted. Grades 1,2,3,4 represent the first grade, the second grade, the third grade, and the fourth grade (and above). However, the reader status, department, and the Most Borrowed Sub-Library are discrete variables, so one-hot is used to transform the data.

This paper directly convert the reader status using 0, 1, 2, but the reader department, and the Most Borrowed Sub-Libraries have discrete variables of 58 and 16, respectively. If the data is directly converted by one hot coding, the final data set will be sparse. Therefore, the variables are classified as follows: the departments are classified into Arts departments, Science and Engineering departments, and Medical departments; the most borrowed sub-library are

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202003_3(3).0022

classified into the Art Sub-library, Science Sub-library, Medical Sub-library. Then one hot coding is carried out for each indictor. For example, the transformation of departments is as follows:

		Arts departments	Science and Engineering departments	Medical departments
Arts departments	One-hot →	1	0	0
Science and Engineering departments		0	1	0
Medical departments		0	0	1

Figure 1. The one-hot transformation of discrete data

2. Determine the cluster center.

Taking the borrowing data of Fudan University Library as the data set, the clustering centers are determined, that is, they are clustered into several categories. It is mainly based on the principle of clustering, that is, the distance between samples within a group is the smallest, and the distance between samples is the largest. When all samples are grouped into one category, the sum of squares of deviations within a group is the largest. With the increase of the number of clusters, the sum of the squares of the deviations in the group will gradually become smaller. When the sum of the squares of the deviations in the group starts from a certain point and the speed of the decline changes from fast to slow, it can be considered that this point is the number of clusters, that is, K cluster centers. In this paper, R language is used to calculate and draw a graph of cluster number selection (as shown in Figure 1). From the figure, it can be known that when the number of clusters is 6, that is, K = 6, the optimal number of clusters. Therefore, 6 samples are selected as the cluster centers.

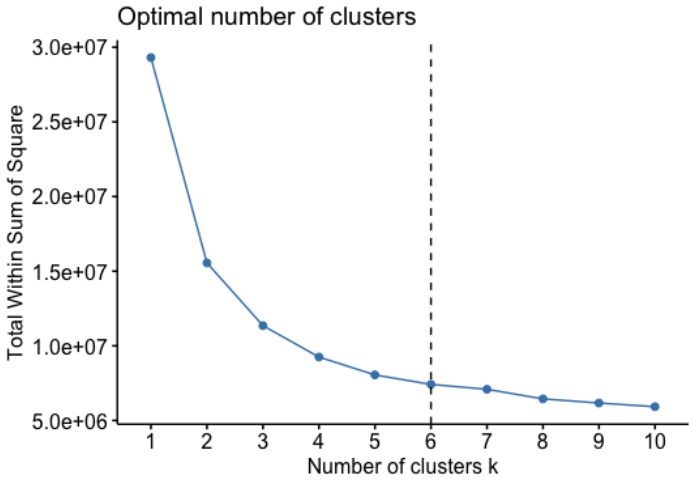

Figure 2. Define cluster center

3. Draw clustering graph and determine the number of clusters.

Based on the six clustering centers, the most similar samples are calculated and visualized. As shown in Figure 2, the borrowing data of Fudan University Library are grouped into six categories, that is, six groups of user profiles. The number of users in each category is as follows: cluster 1: 8912, cluster 2: 236, cluster 3: 3589, cluster 4: 29, cluster 5: 692, cluster 6: 1509.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202003_3(3).0022

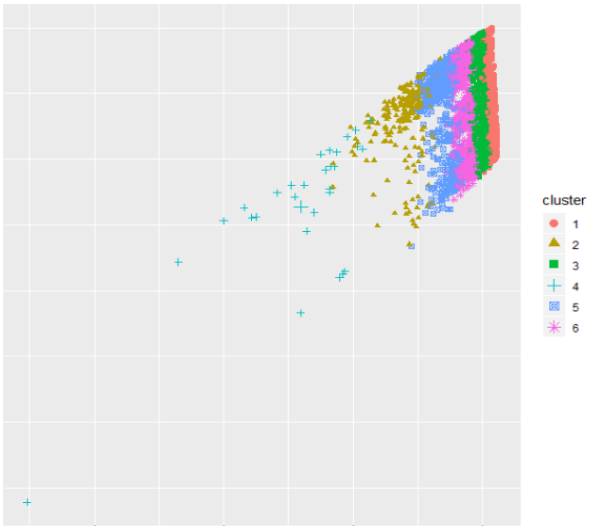

Figure 3. Clustering Graph

5. Analysis of User Portrait Results

According to the clustering graph of readers, the student users of the library are roughly divided into the following six categories:

1. The cluster1 of user profile analysis:

There are 8912 users in cluster1. From the perspective of characters, this cluster users are mainly the first and second year undergraduate students. They come from Marxism college, Journalism college, Zhejiang West Lake Institute of higher learning, and Chinese Language and Literature Department. From the perspective of location, the most borrowed sub-library includes literature of Arts Museum Art Center of Art Sub-library, Political/Economic/Legal of Art sub-library. From the perspective of events, the average number of readers who borrow in the past year are six books, the average renewal are two books, and the average reservation are two books. From the perspective of time, readers tend to borrow at the beginning of the semester, with more in the afternoon. From the perspective of activity, the more active the readers are, the more they borrowed in the past six months. The average borrowing is one book, far lower than the average borrowing in the past year, so this kind of users are very inactive user groups in the library.

2. The cluster2 of user profile analysis:

There are 236 users in cluster2. From the perspective of characters, this cluster users are mainly the students in the second and third year of master's degree and the second year of doctor's degree. They are generally from the school of economics, the school of international relations and public affairs, the school of clinical medicine, the school of social development and public policy, the school of foreign languages and literature, and the department of Chinese language and literature. From the perspective of location, the most borrowed sub-library includes literature of Arts Museum Art Center of Art Sub-library, Political/Economic/Legal of Art sub-library and Medical sub-library. From the perspective of events, the average number of readers who borrow in the past year are 124 books, the average renewal are 30 books, and the average reservation are 20 books. From the perspective of time, readers tend to borrow at the beginning of the semester, with more in the afternoon. From the perspective of activity, the average borrowing are 24 books in the past six months, so this kind of users are more active user groups in libraries.

3. The cluster3 of user profile analysis:

There are 3589 users in the cluster3. From the perspective of characters, this cluster users are mainly the students in the first and second year of undergraduate and master's degree, who

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202003_3(3).0022

come from law school, higher education research institute, department of tourism, Marxism college, department of cultural relics and Museum, literature information center, philosophy college. From the perspective of location, the most borrowed sub-library includes the Arts sub-library of Jiangwan Campus, Science sub-library, and Political/Economic/Legal of Art sub-library. From the perspective of events, the average number of readers who borrow in the past year are 20 books, the average renewal are 5 books, and the average reservation are 4 books. From the perspective of time, readers tend to borrow at the beginning of the semester, with more in the afternoon. From the perspective of activity, the average borrowing are 4 books in the past six months, so this kind of users is the inactive user group of the library.

4. The cluster4 of user profile analysis:

There are 29 users in cluster4. From the perspective of characters, this cluster users are mainly sophomores, who come from the Institute of ancient books arrangement, the school of international relations and public affairs, the research center of historical geography and the department of history. From the perspective of location, the most borrowed sub-library includes literature of Arts Museum Art Center of Art Sub-library, Political/Economic/Legal of Art sub-library. From the perspective of events, the average number of readers who borrow in the past year are 284 books, the average renewal are 15 books, and the average reservation are 37 books. From the perspective of time, readers tend to borrow at the beginning of the semester, with more in the afternoon. From the perspective of activity, the average borrowing are 60 books in the past six months, so this kind of users is very active user group of the library.

5. The cluster5 of user profile analysis:

There are 692 users in cluster5. From the perspective of characters, this cluster users are mainly students in the third and fourth year of undergraduate course and the third year of master's degree. They come from the school of computer science and technology, the school of microelectronics, the department of chemistry, the school of basic medicine, the school of mathematics and the school of management. From the perspective of location, the most borrowed sub-library includes the basic discipline lending room of the science sub-library, the foreign teachers center of the science sub-library, the applied discipline lending room of the science sub-library, the Medical sub-library, and the Zhangjiang Library of the Zhangjiang Campus. From the perspective of events, the average number of readers who borrow in the past year are 66 books, the average renewal are 28 books, and the average reservation are 12 books. From the perspective of time, readers tend to borrow at the beginning of the semester, with more in the afternoon. From the perspective of activity, the average borrowing are 13 books in the past six months, so this kind of users is the active user group of the library.

6. The cluster6 of user profile analysis:

There are 1509 users in the cluster6. From the perspective of characters, this cluster users are mainly sophomores. They are generally from the natural science experimental class, the technical science experimental class, the economic management experimental class, the social science experimental class, the clinical medicine school and the pharmaceutical college. From the perspective of location, the most borrowed sub-library includes the Arts Comprehensive sub-library, literature and Art Center of Arts sub-library, Political/Economic/Legal of Art sub-library, Art sub-library in Jiangwan campus, Chinese book of Medical sub-library. From the perspective of events, the average number of readers who borrow in the past year are 39 books, the average renewal are 12 books, and the average reservation are 8 books. From the perspective of time, readers tend to borrow at the beginning of the semester, with more in the afternoon. From the perspective of activity, the average borrowing are 8 books in the past six months, so this kind of users is the not very active user group of the library.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202003_3(3).0022

6. Analysis and Suggestion

Based on the above analysis of the user clusters, according to the characteristics of the users, in order to improve the knowledge service status of the library and improve the borrowing of books in the library, the following suggestions are put forward for the library:

The library should take measures to carry out a lot of training. For the junior undergraduates, the library should take appropriate measures to improve the reading situation. In theory, as the mainstay of the user group of the library, the number of undergraduates should account for the majority of the total amount of borrowing. However, the current situation of undergraduates in the library is not very optimistic. The junior undergraduates have just stepped into the university, they may not know the library's borrowing, renewal and appointment process, library collection distribution and library related services very well. Therefore, it is necessary for the librarians to take corresponding training methods, so that the junior students get familiar with the library's services and collection as soon as possible, to promote the students to borrow books.

Libraries take targeted measures to implement recommendations for different types of user clusters. Generally, the humanities and social sciences majors need to read a large number of classic books, so the students have a large demand for books. For this kind of user cluster, the library can recommend books, new books and intensive library books that are consistent with the major in real time. For the major of science and engineering and medical science, first of all, we need to have a certain theoretical and knowledge base. For this type of students, we should buy more systematic and theoretical books, and recommend them to the students of the lower grades according to their major, so as to lay a good foundation for their study.

The library should improve its knowledge service mode and make use of big data and artificial intelligence technology to make personalized recommendation. For example, the library can launch a personalized recommendation system for the collection of books. According to the user's behavior data and personal basic information, it can implement one-to-one accurate recommendation for users, on the one hand, improve the working efficiency and service quality of the library, on the other hand, save user time and enhance user experience.

7. Limitation and Future Work

Although based on the actual data of university library as an example for the empirical analysis, but the choice of data type and the amount of data has some limitation, this article only on the basis of borrowing data to build models, in fact, the relevant behavior data, such as: retrieve data, browsing data, collecting data, and so on can be used as the data source user base to build a subdivision profile groups of users, but because of those data cannot be collected , so this research only based on the current available data for research. In the subsequent research intends to select the more rich data types and a greater amount of data, build more detailed user profile, and in-depth analysis of characteristics of different categories of users, and behavior habits, further support more accurate university library books recommended, for academic researchers and university library to provide more theoretical and practical guidance.

8. Conclusion

In the context of big data, in order to improve the utilization rate of books in university libraries, this paper explores the construction of user profile model in university libraries. On the basis of borrowing data on August 31, 2017 solstice and September 1, 2018 in the library of Fudan University, the k-means clustering algorithm in data mining technology is used to divide the group user profile into six categories. Then, the analysis of the user profile was made, and corresponding Suggestions were proposed for the library.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202003_3(3).0022

References

[1] Cooper, Alan, et al. About face: the essentials of interaction design(John Wiley & Sons, 2014).

[2] Qing Wang, Fazhen Zhao. Design and Analysis of Library Resource Recommendation Model Based on User Profile, Journal of Modern Information, Vol. 38(2018), No.3, 105-109+137.

[3] Datian Bi, Fu Wang, Pengcheng Xu. Analyzing Mobile Library Users and Recommending Services with VSM, Data Analysis and Knowledge Discovery,Vol.2(2018),No.9,100-108.

[4] Herrera-Viedma, Enrique. Using Bibliometrics and Fuzzy Linguistic Modeling to Deal with Cold Start in Recommender Systems for Digital Libraries. Advances in Fuzzy Logic and Technology 2017: Proceedings of: EUSFLAT-2017–The 10th Conference of the European Society for Fuzzy Logic and Technology, (September 11-15, 2017, Warsaw, Poland IWIFSGN) Vol. 3.

[5] Jomsri P. Book recommendation system for digital library based on user profiles by using association rule[C]. 4th International Conference on Innovative Computing Technology, INTECH 2014 and 3rd International Conference on Future Generation Communication Technologies, FGCT 2014, 2014:130-134.

[6] Kuzelewska, Urszula. "Clustering algorithms in hybrid recommender system on movielens data." Studies in logic, grammar and rhetoric Vol.37 (2014), p.125-139.

[7] Kovacevic A, Devedzic V, Pocajt V. Using data mining to improve digital library services.Electronic Library, Vol.28(2010), No.6, p.829-843.

[8] Shirude, Snehalata B., and Satish R. Kolhe. Agent-based architecture for developing recommender system in libraries. (Knowledge Computing and its Applications. Singapore, 2018,p.157-181).

[9] Yuan Hu,Ning Mao. User Modeling of Digital Library Knowledge Community Based on User Portrait, Library Theory and Practice.2017, Vol.2(2018), 82-85+97.

[10] Tianyuan, Chen. An Empirical Research on Personas Construction of Mobile Library in Universities, Library and Information Service, Vol. 62(2018) No.7, 38-46.

[11] Fan Yang. Library Big Data Practice Based on User Profile Analysis——A Case Study of National Library of China. Library Tribune.Vol.39(2019) No.2, 58-64.

[12] Juan He. Application Research on Personalized Recommendation of Books Based on User Portrait and Group Portrait. Information Studies:Theory & Application.Vol.42(2019) No.1,p. 129-133, 160

[13] Hailing Xu, Haitao Zhang, et al. Group User Interests Profile in University Libraries Based on Concept Lattice, Information Science.Vol.37(2019) No.9,p: 153-158+176.