

Association for Information Systems

## AIS Electronic Library (AISeL)

---

ICEB 2015 Proceedings

International Conference on Electronic Business  
(ICEB)

---

Winter 12-6-2015

### Research On The Construction Of Personas Model Based On K-Means Clustering Algorithm

Cheng Quan

Dengfu Liu

Lanfang Zhou

Follow this and additional works at: <https://aisel.aisnet.org/iceb2015>

---

This material is brought to you by the International Conference on Electronic Business (ICEB) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICEB 2015 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# RESEARCH ON THE CONSTRUCTION OF PERSONAS MODEL BASED ON K-MEANS CLUSTERING ALGORITHM

Cheng Quan, School of Economics and Management, Fuzhou University, Fuzhou, Fujian 350116  
Liu Dengfu, School of Economics and Management, Fuzhou University, Fuzhou, Fujian 350116  
Zhou Lanfang, School of Economics and Management, Fuzhou University, Fuzhou, Fujian 350116

## ABSTRACT

As a model framework of outlining the network target group and locking user's demand trait, Personas can be used to reveal the characteristics and behaviors of the user efficiently and accurately. Setting the data of SINA Weibo users as research object, a comprehensive description of the multi-dimensional user characteristics as the main purpose, the K-means clustering algorithm as a technical means, this paper mines the user data deeply, then builds a Personas model, finally visualizes the results of Personas model based on Matlab. In this paper, the proposed method of building a Personas model provides a reference for practical application problems, such as the accurate locking of potential users and the targeted delivery of network advertising.

*Keywords:* K-means algorithm; Personas; High-precision classification

## INTRODUCTION

In network economic era, the rapid development of "internet + trade" will drive user demand for personalized service underlined, and the main mission of network marketing becomes more and more focus on capturing the features of the user accurately. Customer analysis has been an important discussion of commodity sales area, especially under the network environment, customer analysis is influenced by statistic customer descriptive data and dynamic customer behavioral data, and consequently it has distinct features of big data analysis. User experience authority expert Alan Cooper for the first time to put forward the concept of Personas [4]. Anhua Ma [1] divided Personas dimensions into static information and dynamic information in the design and implementation of precision marketing system, classified the required data of building Personas model; Mengjie Yu [3] makes use of statistical analysis, specific qualitative individual description and statistical thinking to put forward in product development to build Personas model; Shuguang Zhao [6] through in-depth interviews with 500 users of social media use motivation and behavior, divided social users into five categories by quantitative statistical analysis, and formed a rough Personas model. The Personas is a description of the specific details of the target user and a virtual representation of the user, which is based on a series of marketing data, usability data and other real data to build the target Personas model.

Setting the data of Sina Weibo users as research object, a comprehensive description of the multidimensional user characteristics as the main purpose, the K-means clustering algorithm as a technical means, this paper mines the user data deeply, then builds a Personas model, finally visualizes the results of Personas model based on Matlab, hope to be able to inspire the above questions.

## DATA ACQUISITION

Scrapy [5] is a distributed crawler framework based on Python. It provides the abstract design of crawler, which can be easily and efficiently to crawl and analyze all kinds of data on the network. According to the attribute of the Weibo user to define item, spider, item pipeline and other items, access to Sina Weibo API interface, we can start data scrawling.

## K-MEANS CLUSTERING ANALYSIS FOR DATASETS

Through the preprocessing for the original data, the final datasets includes 32,915 user data. For the data is scattered, it is unable to get the relationship and distinction between users. If clustering analysis [2] can be made of the user, just look at a certain class to know the feature of all the users within the class. In the case of no-class label information, clustering analysis groups things automatically to enable every group to identify themselves and distinguish from other groups, so this study makes clustering analysis of the user firstly.

The Andrews<sup>1</sup> figure is used to show clustering results. As shown in Figure 1, every ribbon is on behalf of a class of user, and the wider the ribbon is, the more users is represented. Besides in figure 4, the t axis represents the observed value, and the f (t) axis represents a function of t.

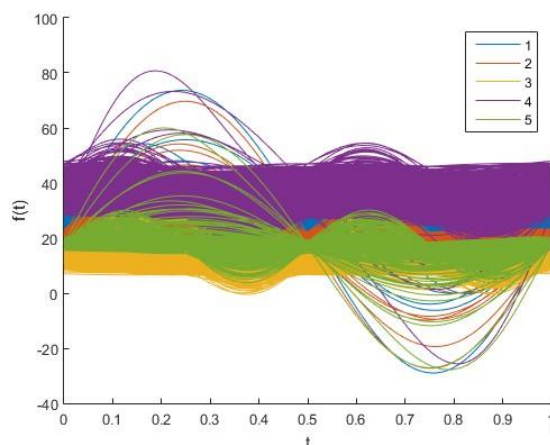


Figure 1. The Effect Diagram of K-means User Clustering

From the effect diagram of K-means user clustering, the users whose data is centralized are showing five kinds of features:  
 The first kind of users: male users who are about 35 years old have quite low influence and active degree, and are from non-mobile platform.

The second kind of users: young female users who are 28 years old have quite high influence and active degree, and are from mobile platform.

The third kind of users: young male users who are 19 years old have low influence and active degree, and are from non-mobile platform.

The fourth kind of users: middle-aged male users who are 45 years old have very high influence and active degree, and are from non-mobile platform. This kind of users is the opinion leaders of Weibo.

The fifth kind of users: young female users who are 22 years old have general influence and active degree, and are from mobile platform.

Up to this point, preliminary clustering of Weibo user is finished.

### USER LABEL ANALYSIS

The user label is a static data that can represent the user's personal characteristics. This study will be based on the five categories of user labels that are classified by K-means clustering, and then obtain the information to improve the user's characterization. Specific processing steps are as follows:

(1) a label for each user (a single user can have multiple labels) in a two-dimensional table, each row in a table represents a user's label.

(2) for the user's label, because the label is too mixed, the research will be carried out by the user label and delete the independent label, a total of 8 categories of labels. Each user's label is divided into the 8 categories according to the subject characteristics, and the number of users per class label.

(3) through the data perspective table, 32915 users of each user's label categories and their number, and then combined with the data analysis of the 5 categories of clustering results, 5 types of basic user's label categories and specific numbers, as shown in table 1:

Table 1. User Label Statistics

Label	1st	2nd	3rd	4th	5th
Food	492	1247	1723	313	1790
Film	141	511	1175	27	911
Sports	131	349	445	64	590
Games	0	133	451	0	339
Digital	261	492	550	119	782
Car	153	274	304	69	411
Dress	352	1339	1484	288	1740

Music	316	1094	1964	287	1648
-------	-----	------	------	-----	------

By Table 1, the number of "music" labels is far more than the rest of the labels, and then dress and food. Although the multi label is helpful to accurately portray the microblogging users, but the post ranking label has a low impact, too many labels will also give users a picture after the application inconvenience. Therefore, on this basis, this study selects the top four of the label to reveal the user portrait.

Such as the first category of user groups, the top four of its users for food, clothing, music and digital, you can describe the type of user attention fashion, focus on quality of life of a class of people. The rest of the 4 categories can be described in accordance with this idea of the user, specifically see table 2:

Table 2. user Labels

Label	1st	2nd	3rd	4th	5th
Label 1	Food	Dress	Music	Food	Food
Label 2	Dress	Food	Food	Dress	Dress
Label 3	Music	Music	Dress	Music	Music
Label 4	Digital	Film	Film	Digital	Film

### USER BEHAVIOR ANALYSIS

After the user's static characteristics extracted from the user's labels, then the analyzed user's behavior data to get user dynamic features.

In this study, we use the data perspective to analyze the regional distribution of the Weibo users, as shown in Figure 2:

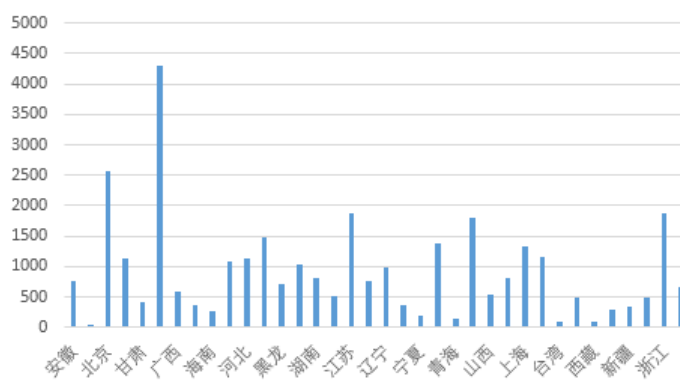


Figure 2. Weibo User Area Map

All kinds of user area distribution map is as follows:

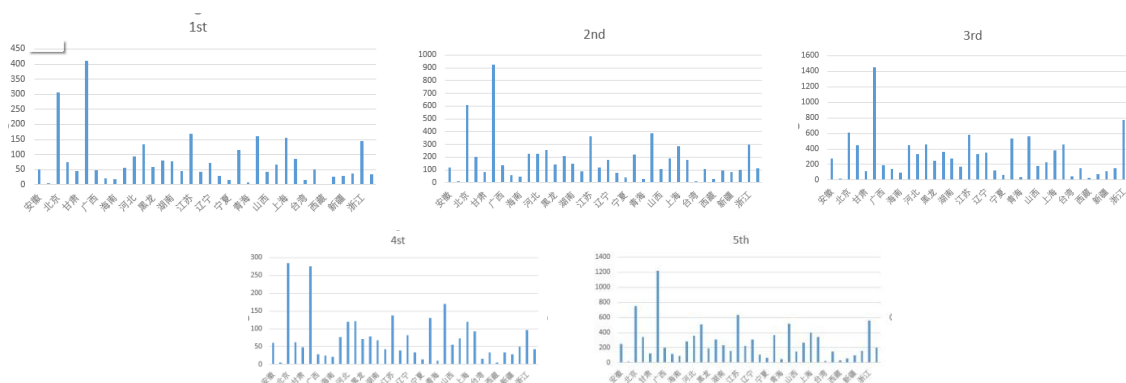


Figure 3. Five types of user area distribution map

It can be seen from the picture above that in the wider region of Guangzhou, Shanghai and Beijing Weibo users are more concentrated, these regions are densely populated areas. The rest of the regional Weibo users basically uniform distribution, the deviation is not large.

Secondly, for the analysis of the Weibo user search keywords, the same as user label analysis, the study extract and statistics keywords. Use frequency in the first three words as the user behavior features, as shown in the following table:

Table 3. User Keywords

Keywords	1st	2nd	3rd	4th	5th
Keyword 1	Music	Clothes	Computer	History	Snacks
Keyword 2	Education	iphone	Food	Book	Li Jian
Keyword 3	Xiaomi	Travel	Guitar	Western-style Clothes	Meizu

### PERSONAS MODEL CONSTRUCTION

Through K-means clustering algorithm is used for static feature extraction and dynamic behavior mining to user data above, the Personas formed has the data support and the rich, and meet the basic requirements of the Personas principles by Alan Cooper basically. Sina Weibo users whose raw data is centralized are divided into five categories of user groups subtly:

The first category of users: male users who have quite low influence and activeness age in 31 and 35 years old, and use non-mobile platform; individual label category: Food, Dress, Music, Digital; search keywords category: Music, Education, Xiaomi.

The second category of users: young female users who are not the opinion leaders of Weibo have quite high influence and activeness, age in 26 and 30 years old, use mobile platform; individual label category: Dress, Food, Music, Digital; search keywords category: Clothing, iPhone, Travel.

The third category of users: young male users who have low influence and activeness age in 16 and 20 years old, and use non-mobile platform; individual label category: Music, Food, Dress, Film; search keywords category: Computer, Food, Guitar.

The fourth category of users: middle-aged male users who are the opinion leaders of Weibo have very high influence and activeness, age in 41 and 45 years old, use non-mobile platform; individual label category: Food, Dress, Music, Digital; search keywords category: History, Book, Western-style Clothes.

The fifth category of users: young female users who have general influence and activeness in 21 and 25 years old, and use mobile platform; individual label category: Food, Dress, Music, Film; search keywords category: Snack, Li Jian, Meizu.

Visual performance of these five classes of users is shown in figure 4:



Figure 4. Sina Weibo user Personas Model

### CONCLUSION

Personas model getting through the above process can accurately classify the user, describe the user features in detail, solve the

problems arising from the user classification before. It is a better user positioning method. Broadly speaking, any need to locate users precisely or virtual users is able to use personas model to realize. Building network personas model provides a necessary reference for accurately delivering network marketing.

#### REFERENCES

- [1] Anhua Ma.(2013) *Design and realization of accurate selling system based on behavior analysis*, Nanjing University of Posts and Telecommunications.
- [2] JiaweiHan, MichelineKamber, Fan Ming & Mengxiaofeng (2007) *Data mining: Concepts and techniques*, Beijing: China Machine Press.
- [3] Mengjie Yu.(2014) 'Data modeling of user portrayal in product development: From concrete to abstract', *Design Act Research*, No. 6, pp.60-64.
- [4] Personas, available at <http://www.usability.gov/how-to-and-tools/methods/personas.html>.
- [5] Scrapy, available at <http://www.scrapy.org>.
- [6] Shuguang Zhao.(2014) 'High conversion of social media users portrait: Based on the deep interview of 500 users', *Modern communication (Journal of Communication University of China)*, No. 6, pp. 115-120.