Research on Categorization of Animation Effect Based on Data Mining

Na Ni

Department of Animation, Art College, Henan University, Kaifeng, Henan, China

ABSTRACT: Nowadays, the production process of animation effect is increasingly developed, and its effect is also growing better. But in most cases, the categorization of special effect added to the animation is confusing due to excessive variations. Data mining will desirably solve the problem of animation effect categorization, so the application of data mining in the animation effect categorization becomes the hot spot in research and analysis at present. This article makes a detailed analysis on relevant algorithm of data mining technology, that is, the k application of averaging method, k central point method and relational degree algorithm in problem of animation effect categorization. It provides a clear method of categorization for animation effect. Thereafter, it also concludes the accuracy of animation effect categorization can be greatly improved through reasonable algorithm integration in the treatment of animation effect categorization by data mining.

Keywords: data mining, animation effect categorization, cluster analysis, relational degree

1 INTRODUCTION

In the background of opening up policy, the establishment of socialist road with Chinese features realizes rapid development for various trades in China. The incorporation of diversified elements and various industries provides unique vitality for various industries, and our Chinese production process of animation effect also rapidly develops from immaturity to excellence. We are always learning and trying to use the globally advanced new technologies. Prepare the whole animation effect based on thinking and experience in previous animation making, different eras, objectives and prominent effect. So there are many kinds of animation effect with different advantageous functions. Therefore, how to distinguish a great number and many types of animation effect and provide obvious categorization index becomes the hot spot nowadays. In the current big data era, many use the mathematic method to solve the actual problems, and the methods are increasingly perfected. This article tries to use data mining method for detailed research and analysis on the problem of animation effect categorization.

Figure 1. Animation effect of hierarchy sense
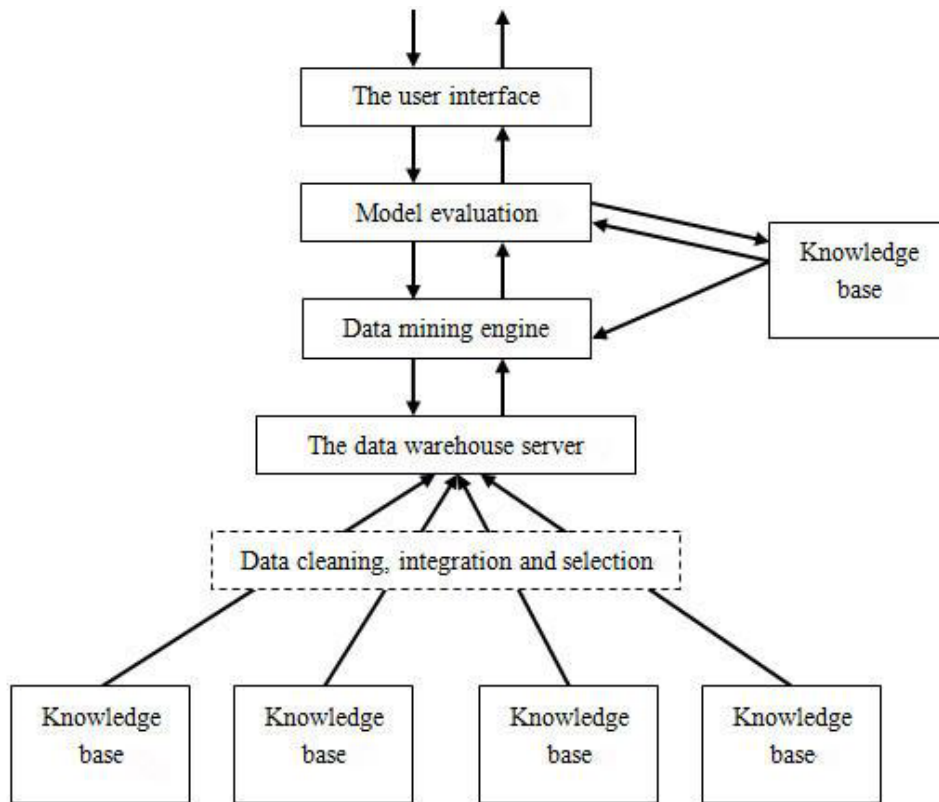
Figure 2. Animation effect of fuzzy

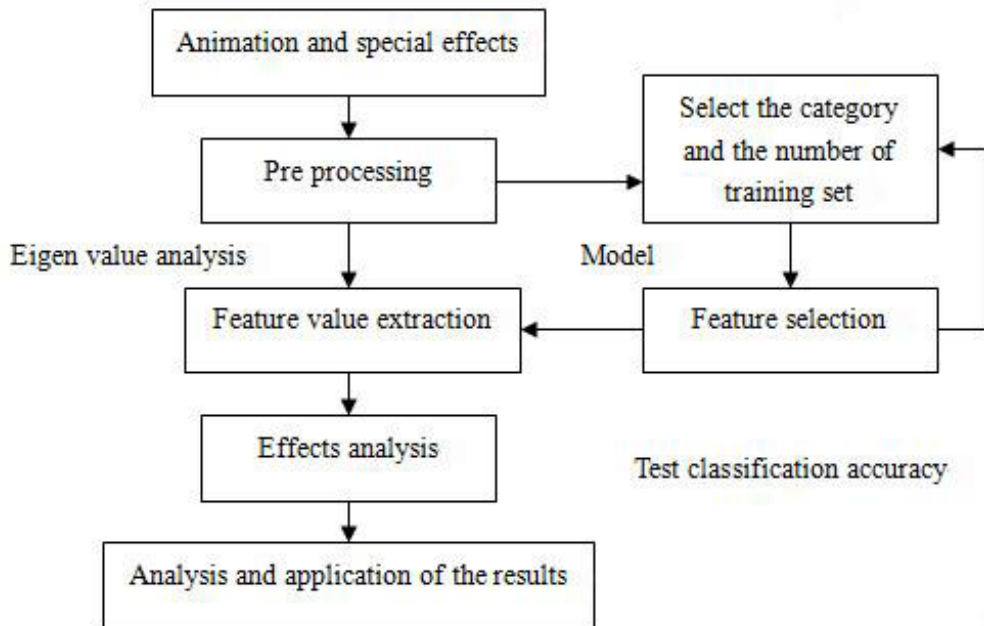Figure 3. Typical flow chart of data mining

Figure 4. Flow chart of data mining for animation effect categorization

2 MODEL ESTABLISHMENT AND ITS SOLUTION

In the current society with developed mass media, there are increasingly growing types of animation effect that are used to improve the animation effect and have more comprehensive functions. For example, the animation with many scenic spots will require special effect to strengthen the sense of hierarchy, and the animation requiring artistic conception shall be added with special effect of local fuzzy or rendering; the action scene with visual impact requires improved special effect of textile, and so on. Figure 1 and Figure 2 add the special effect of hierarchy sense and fuzzy for the animation to provide satisfactory visual effect for the animation.

It can be seen from Figure 1 and Figure 2 that, despite the emphasis of those two Figures, in most cases the animation is composed by multiple special effects, so the special effect categorization added for the animation becomes fuzzy and the accuracy will be greatly reduced. Data mining will desirably solve the problem of animation effect categorization, so the application of data mining in the animation effect categorization becomes the hot spot in research and analysis at present.

2.1 *Principle of data mining for animation effect categorization*

The data mining refers to extract the hidden information which is hard to be identified by human efforts by mining and collecting a large number of fuzzy data or partial data; then structure and screen out available information therefrom; and deduct, generalize and make statistics; then finally visualize those information, and find out the relevance of those data by macro analysis and research; categorize them based on comprehensive consideration of their typical features. While the data mining algorithm used for treatment of animation effect categorization problem in this article mainly focuses on extracting and mining the types and special effect of production process for each animation effect, generalizing and making statistics therefor, finding out the relevance of various animation effects and finally categorize and identify them based on relevance of various animation effects. Figure 3 is the typical flow chart of data mining.

It is known from the figure 3 that the traditional data mining method is to use various data analysis algorithm based on the obtained data, to establish a model, to categorize and assess the obtained data; predict the variables and finally to concentrate and group according to the relevance. While for the treatment of animation effect categorization, it mainly mines and analyzes the 3D or multi-dimensional data. Figure 4 is the flow chart of data mining for animation effect categorization.

It can be seen from Figure 4 that this article makes pre-treatment for the animation effect; invests it into type and training concentration; then comes into feature selection mode; conducts feature analysis for obtained data; extracts feature data; categorizes the animation effect and inspects the categorization accu-

racy of obtained result. Thus it will be avoided to select the data without differences and will make other serious mistakes. And the final inspection of categorization accuracy for animation effect categorization is also an important index for assessment model accuracy, and it imposes a great impact on the clarity and accuracy of animation effect.

2.2 *Analysis on data mining algorithm*

As mentioned above, the data mining refers to extracting the hidden information hard to be identified by human efforts by mining and collecting a large number of fuzzy data or partial data, then structuring and screening out available information therefrom, and making mathematical statistics, then analyzing the relevance of those data, and finally categorizing them based on comprehensive consideration of their typical features. This article uses three data mining algorithms to conduct corresponding analysis and research respectively on the problems of animation effect categorization.

2.2.1 *Analysis on cluster algorithm*
The cluster analysis is a process that generalizes and categorizes the obtained data according to its similarity, and put the similar data into one set. Then mark every obtained data aggregate. This cluster algorithm analysis may solve variant problems and desirably find the relative features for various data aggregate. There are two types of data aggregate in the cluster analysis, which are data matrix and adjacent value matrix. Data matrix, that is, the matrix of $m \times n$ by representing n target with m variables, shown as follows:

$$\begin{bmatrix} x_{11} & L & x_{1f} & L & x_{1p} \\ L & L & L & L & L \\ x_{i1} & L & x_{if} & L & x_{ip} \\ L & L & L & L & L \\ x_{n1} & L & x_{nf} & L & x_{np} \end{bmatrix}$$

While adjacent value matrix refers to a matrix of $n \times n$, which represents n matrix of data similarity as follows:

$$\begin{bmatrix} 0 & & & & \\ d(2,1) & 0 & & & \\ d(3,1) & d(3,2) & 0 & & \\ L & L & L & 0 & \\ d(n,1) & d(n,2) & L & L & 0 \end{bmatrix}$$

Among which, the similarity measured between data i and data j is $d(i,j)$. As $d(i,j)$ approaches 0, it indicates the data i and j are more similar; similarly, as $d(i,j)$ value grows, it indicates the data i and j have less similarity.

So there may be errors between data aggregates in different categorizations. While the error calculation method used in the cluster analysis of this article is as follows:
1) The absolute error of mean:

$$S_f = \frac{1}{n}\left(\left| x_{1f} - m_f \right| + \left| x_{2f} - m_f \right| + L + \left| x_{nf} - m_f \right| \right)$$

x_{1f}, L, x_{nf} represent n measured values for

f, while the average of f is m_f, and

$$m_f = \frac{1}{n}\left(x_{1f} + x_{2f} + L + x_{nf} \right)$$

2) Standard measured value:
The absolute error of mean proposed above can desirably reduce the impact of outliers on calculation result. That is to say, the measured value of outliers obtained from absolute error of mean can be well identified. The calculation formula with measured value is as follows:

$$z_{if} = \frac{x_{if} - m_f}{s_f}$$

The similarity between data aggregates in cluster analysis is determined according to comparison of adjacent values for all data. While there are mainly two calculation methods for $d(i,j)$:
1) Euclidean distance:

$$d(i,j) = \sqrt{\left(x_{i1} - x_{j1} \right)^2 + \left(x_{i2} - x_{j2} \right)^2 + L + \left(x_{in} - x_{jn} \right)^2}$$

$$i = (x_{i1}, x_{i2}, L, x_{in}) \qquad \text{and}$$

$j = (x_{j1}, x_{j2}, L, x_{jn})$ in the formula above respectively represents the data in two n-dimensional data aggregate.
2) Manhattan distance:

$$d(i,j) = \left| x_{i1} - x_{j1} \right| + \left| x_{i2} - x_{j2} \right| + L + \left| x_{in} - x_{jn} \right|$$

The Euclidean distance and Manhattan distance above shall meet the requirements below, namely:

$$\left\{ \begin{array}{c} d(i,j) \geq 0 \\ d(i,i) = 0 \\ d(i,j) = d(j,i) \\ d(i,j) \leq d(i,h) + d(h,j) \end{array} \right.$$

The cluster algorithm analysis introduced in this article solves the categorization problem of animation effect with two methods, namely k averaging method and k central point method. The k averaging method is to divide n data into k data aggregates by taking k as the parameter, and the categorization result is the higher data similarity in the data aggregate which leads to higher difference between data aggregates. However, the similarity of data aggregate is its barycenter. It is defined as follows:

$$E = \sum_{i=1}^{k} \sum_{p \in C_i} |p - m_i|^2$$

E represents the sum of quadratic difference of all data in the data aggregate, p the specified data point, and m_i the mean of data aggregate C_i. Randomly select k data in the data set D as the initial center of data aggregate, judge each data according to the formula above through the obtained mean, put it into the data aggregate with the most similar features, change the mean of data aggregate, and calculate the mean of data until it does not change. Therefore, the k averaging method will have a better performance when the data aggregate separated from animation effect has a relatively obvious difference. While the k central point method is to select a data in each data aggregate to represent the whole data aggregate, and cluster the rest of data into the data aggregate of selected data with the highest similarity in order to categorize the data of animation effect. It is defined as follows:

$$E = \sum_{i=1}^{k} \sum_{p \in C_i} |p - o_j|$$

o_j represents the data selected from C_j. Now we randomly select k data from data set D as the initial selected data; assign the rest data into the data aggregate represented by the selected data according to the above formula; select one more data o; change o_j into o; if the generally paid function is $S < 0$, a new data set will be formed, and the repeated calculation is required until it is unchanged. The result calculated by central point is used to assess the mean of dissimilarity between selected and residual data through the function S, but there are four cases as to whether the data o_j can be changed into o.
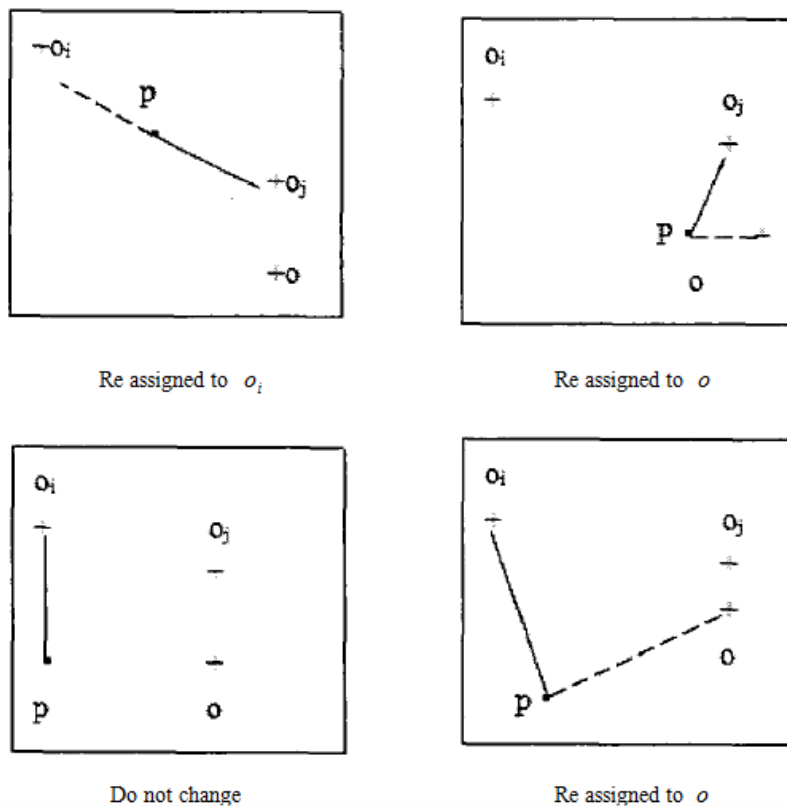
Figure 5. Four statuses to change non-representative data

As known from the figure above, when p is in the data aggregate represented by o_j, if o_j is changed into o, and p is closest to the residual selected data o_i, p shall be put into o_i; when p is in the data aggregate represented by o_j, if o_j is changed into o, and p is closest to the data o, p shall be put into o; when p is in the data aggregate represented by o_i, if o_j is changed into o, and p is closest to the data o_i, it shall not be subject to a second assignment; when p is in the data aggregate represented by o_i, if o_j is changed into o, and p is closest to the data o, the p shall be put into o; every assignment will make the S function of E change. If $S < 0$, a new data set will be formed; If $S > 0$, it will indicate the change can be accepted, and thus it is not subject to another assignment, that is to say the categorization is finished.

2.2.2 Analysis on relational degree algorithm

Relational degree algorithm refers to an algorithm that can desirably find out the relationship and relational degree between obtained data sets after obtaining the valid data through various data mining methods and algorithms. Such as $A = a \Rightarrow B$, that is to say, when the feature value of A is a, the feature value of frequent item set B will have a higher probability to be b. So the relational degree algorithm may be used to detect the abnormal data and achieve high accuracy with the help of other cluster algorithm.

To find out relational degree, we shall first obtain the frequent item set and the relevance among the data. And the frequent item set refers to elements with high appearance rate in the data aggregate, and its threshold scope is defined by support function. Its function is as follows:

$$Support(A, B) = P(AYB)$$

$$Support(\{A, B\}) = \frac{N_{(A,B)}}{N}$$

Among which, N is the total number of times, and $N_{(A,B)}$ the number of times the elements A and B appear simultaneously. As the minimum support threshold Min-sup is found, this parameter may only capture the data larger than Min-sup, so it will be easy to obtain the frequent item set in the objective aggregate. As the frequent item set is found, we may calculate the relational degree, that is to say we may get a relevance that has higher relational degree and an important impact on the analysis. When calculating the relational degree, its confidence level shall be met. The calculate formula is as follows:

$$confidence(A \Rightarrow B) = P(B|A)$$

$$confidence(A \Rightarrow B) = \frac{Support(\{A, B\})}{Support(\{A\})}$$

The formula above indicates the confidence level of $A \Rightarrow B$ is the probability of B under A condition, namely the confidence level of $A \Rightarrow B$ may be obtained through calculation of $Support(\{A, B\})$ and $Support(\{A\})$, which also indicates $A \Rightarrow B$ can be easily calculated after determining the value of $Support(\{A, B\})$ and $Support(\{A\})$, that is to say we may obtain the corresponding relational degree among data and inspect its accuracy after determining the value of $Support(\{A, B\})$ and $Support(\{A\})$.

It is known from the analysis above that, in solving the problem of animation effect categorization, the data mining identifies and categorizes the special effect based on analysis of creation purpose and process of animation effect, as well as through discovering and extracting the data of animation effect. This article categorizes the animation effect with k averaging method and k central point method and inspects the abnormal data of animation effect categorization by incorporating k central point method and relational degree algorithm to improve the categorization accuracy.

3 CONCLUSION

This article will first analyze the development situation of current animation effect according to the detailed situation of society, and then find out the main reason that causes the problem of animation effect categorization as well as the necessity to apply the data mining therein. This article makes a detailed analysis on relevant algorithm of data mining technology, namely the application of k averaging method, k central point method and relational degree algorithm in problem of animation effect categorization. It provides a clear method of categorization for animation effect.

In solving the problem of animation effect categorization, this article uses the data mining to identify and categorize the special effect based on analysis of creation purpose and process of animation effect, as well as through discovering and extracting the data of animation effect. This article categorizes the animation effect with k averaging method and k central point method and inspects the abnormal data of animation effect categorization by incorporating k central point method and relational degree algorithm to improve the categorization accuracy.

REFERENCES

[1] Faber, L. & Walters, H. 2004. *Animation Unlimited*. Translated by Wang, K.Y. Shanghai People's Fine Arts Publishing House.

[2] Stephen Visiro. 2004. Digital Cartoon Art. GU, C. Shanghai People's Fine Arts Publishing House.

[3] Steve Caplin, Adam, Banks. 2004. Digital Graphics Art. Translated by Cao, T.Q. Shanghai People's Fine Arts Publishing House.

[4] Liu, F., Wu, P... Abnormal Data Capture and Analysis Based on Data Mining. The research China information system.

[5] Qi, H.Y., Wu, X.J., Wang, S. & Yang, J.Y. 2010. A Collaborative FCPM Fuzzy Clustering Algorithm, *Pattern Recognition and Artificial Intelligence*, 23 (1): 120-122.

[6] Gan, L. & Wang, X. 2008. Study on Cluster Analysis in Data Mining. *Coastal Enterprises and Science And Technology*, 42 (3): 24-26.

[7] Liu, H.Y., Chen, J. &Chen, G.Q. 2002. Review of Data Categorization Algorithms for Data Mining. Journal of Tsinghua University (NATURAL SCIENCE EDITION), 42 (06): 727-730.

[8] Kong, L.F., Wang, H. & Lian, Q.S. 2006. Data Organization Method Based on Association Rules Mining. *Computer Engineering*, 32 (21): 12-15.

[9] Fu, Z.W... 2010. Research on Incremental Updating Algorithm for Association Rules. Journal of Software, 9 (8): 52-53.

[10] Li, J.F., Xiao, R. & Qin, J. 2010. Research and Design of Incremental Mining Association Rules. *Journal of Software*, 9 (10): 77-79.