Vast Amounts of Video Data Clean Algorithm Base on Bradley-Terry Model

Binwen Cao[1, 2, a], Li Chen[1, 2, b] and Hui Lie[3, c]

[1]College of Computer Science and Technology, Wuhan University of Science and Technology, Hubei Province

[2]Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan, 430065, Hubei Province

[3]Wuhan Eastern Wisdom Technology Co. Ltd, Wuhan, 430065, Hubei Province

[a]cbw19920506@163.com, [b]chenli@ieee.org, [c]liehui@eastwit.com.cn

Abstract. Traditional video data cleaning method is to use video quality subjective labeling to achieve. This method relies on the subjective and objective factors such as the level of expert knowledge of the observer and the observing environment, so that a high accuracy rate cannot be ensured, and in the massive data conditions need to consume a lot of manpower and time. In this paper, Bradley-Terry model is applied to the field of video cleaning for the first time, let each observer will randomly selected two videos from the sample than Choose the better of the two according to their quality, then we use the ELO Rating System points the results, finally an ordered set is obtained. We use the real-time monitoring of video data to complete the experiment and evaluate our algorithm with stability and effectiveness. The results show that the proposed video quality sorting method has good stability and accuracy, and can be finely and quickly sort the data by their quality in the video big data environment. Experimental results show that the algorithm can clean the video data accurately and quickly under the condition of massive video.

Introduction

Video surveillance is an important part of social public safety. In recent years, with popularity and development of video surveillance systems, more and more industries and units have built a large number of video surveillance systems, using video surveillance to carry out supervision, management and investigation has become an important method. Nowadays, the monitoring system of each profession is more and more huge, the camera and video equipment are more and more. But how to manage and maintain so many devices has become the customer's primary problem [1]. Monitoring video quality is a measure of whether a device is operating properly; therefore, monitoring video quality evaluation has become a research hotspot in the security operation and maintenance industry. But how to monitor the real-time from the massive video data to clean out the structure of the data has not received enough attention. We need knowledge of specific application areas to perform the data cleaning process well, so people need to domain knowledge in the form of rules expressed [2]. Common video data cleaning is mainly divided into labeling and sorting.

Annotation refers to the viewer according to the content of video images to describe the video, For example: Fei-Fei Li who is in the United States Stanford University hosted the Image Net which is based on collecting pictures on the Internet and annotate them, finally let the machine to learn these annotation information so that the computer can identify the contents of the picture, This is a standard logistic regression model. Sorting is a more detailed annotation which is sorting the video images according to specific rules. In this paper we invokes a number of observers to sort the mass of the actual monitored video and then use the data for subsequent video quality evaluation studies, this is a linear regression model. Conventional video quality ranking is achieved through subjective assessment of video quality.

The subjective assessment of video quality is based on the ITU-R Recommendation ITU-R[3] and Recommendation ITU-T P.910[4] which is the subjective video quality assessment method for multimedia applications developed by ITU-T, then let viewers to view the video in a series of controlled circumstances and score, finally Calculate the average opinion points (Men Opinion Score, MOS).The key difference between various subjective video quality assessment methods is the use of single stimuli or double stimuli[5].The double stimulus let the viewer needs to watch the original reference video before scoring each video, and the single stimulus is to view a single video at a time and rate its quality. In the process of subjective labeling, the observer's knowledge level and the difficulty of labeling the video are taken into account to get a new evaluation score to improve the accuracy rate[6].But subjective quality assessment of the experimental process for multiple video objects need to repeat the experiment several times, In the case of massive video the entire evaluation process is time-consuming, expensive and difficult to operate, and There may be discrepancies in the scores due to differences in visual perception and understanding of the video quality of the observer. And the traditional 5-point, 9-point and 11-point system is a discrete rating and too little classification cannot accurately deal with massive video sorting problem. Therefore, this paper presents a video sequence sorting algorithm based on Bradley-Terry Model.

Algorithm and Implementation

The traditional video cleaning using video quality subjective labeling method to complete, But the subjective annotation method is done by the observer scoring each video, so the traditional methods are more dependent on the professional subjective experience and expertise in the field. The video quality is determined by a number of indicators together and it is difficult for an observer to directly give the ranking of the video sample in this set, but the observer can more easily select a better one from a pair of samples. If the Bradley-Terry model [7] is used, statistical methods that rely solely on objective criteria of the data can be established.

In the Bradley-Terry model, if a team or player repeats a match-by-two game, the probability of winning or losing between them can be predicted in the model. If $\prod_{ab}$ represents the probability of a victory over b, then for any one game can be derived $\prod_{ab} + \prod_{ba} = 1$, The Bradley-Terry model can be expressed as:

$$\prod_{ab} = \exp(\beta_a) / [\exp(\beta_a) + \exp(\beta_b)] \tag{1}$$

$\exp(\beta_a)$ Represents the level of individual a, in the model it is called strength. Therefore, when $\beta_a = \beta_b$, $\prod_{ab} = \frac{1}{2}$; when $\beta_a > \beta_b$, $\prod_{ab} > \frac{1}{2}$.

The ELO algorithm [8], which is now commonly used in two-match games, can be thought of as a simplification of the Bradley-Terry model, if we set the current integral of each individual to R_i, its intensity can be transformed by the following equation:

$$\gamma_i = 10^{\log \frac{E_i}{400}} \tag{2}$$

We can see that the current strength of the individual is determined by its current integral.

If the algorithm is applied to video cleaning, we can assume that video the expert of A is marked by the observer as having an abnormal level greater than video B is a function of the difference between the two video exception levels [9]. And the update of the current anomaly level of the video individual is proportional to the deviation of the actual annotation result from the expected result. The specific formula is as follows: If video A and video B current abnormal level is divided into R_A and R_B, the probability that the video A is marked in the comparison can be expressed as:

$$E_A = \frac{1}{1+10^{(R_B - R_A)/400}}$$ (3)

After labeling the observer, the rating of the video is updated according to the game result R_A and the expected value E_A, the specific formula is:

$$R = R_A + K(S_A - E_A)$$ (4)

The process of hierarchical division towards real points is called the convergence process, if the video is divided into the initial R_0, the real score of R_1, the initial video from the R_0 convergence to R_1 be observed a total of N times, and then the convergence rate of the video V can be expressed as:

$$V = (R_1 - R_0)/N$$ (5)

In order to speed up the convergence rate, we need a reasonable choice of K value and video of the initial rating points. According to USCF (American Chess Federation) point's rules, Experiment set the initial level of all video is divided into 2000, and the value of K is determined by the range of rating points. The specific rules are listed in Table 1.

Table 1　K value of the rules

Rank Score R	K Value
$R < 2100$	32
$2100 \le R < 2400$	24
$R \ge 2400$	16

In fact, in the process of subjective annotation, the contrast level of two contrasting videos affects the result; the probability that a video with a high level of abnormality is marked will be higher. But due to the observer's own situation, the attention of the observer will change with time and the limit the level of their professional knowledge will change the result. This is like the player's competitive state by the time and the impact of the external environment. The ELO algorithm simplifies the results to only three factors: 1. before the points; 2. opponents of the points; 3. the results of the game. With the number of matches increasing, each video's abnormal level points will converge toward their true level point; finally, each video can find its place in the collection.

In the experiment, we select a set of abnormal types of video from the actual monitoring video library, and initialize each video grade to 2000. The specific labeling process shown in Fig.1:

①　We put 1000 video data into the matching queue, then we select a video from the matching queue as the contrasting video A and randomly selects one of the video sets with the probability of winning $[0.5\text{-}\tau, 0.5\text{+}\tau], \tau \in [0.01, 0.03]$ compared with the video A which is calculated by the Eq.1 from the matching queue as a contrasting video B.

②　The two videos are manually annotated by the observer and then update the rating of the two videos according to Eq.3, and we remove the two video from matching queues and put them to the waiting queue. This operation is repeated until either the matching queue is empty or the remaining one, then the round ends.

③　After a number of rounds, you can get this group of video based on the level of the exception type ranking. Then compare results with the traditional manual labeling method.
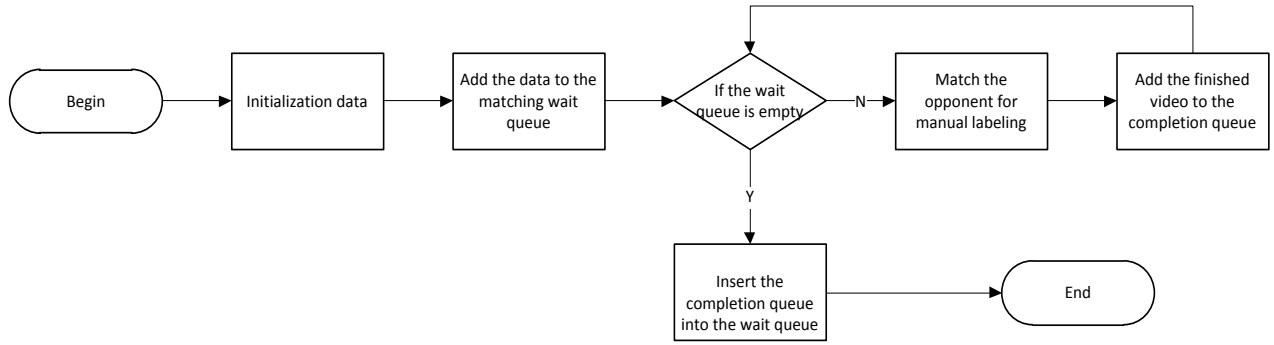
Figure 1.   System operation flow chart

Experimental Results and Analysis

First, we need to conduct a simulation experiment to determine the number of times each video is compared in the Brand-Terry model, and then we selected 2000 videos which were used in this algorithm and MOS score method for experiment. Finally, we compare the efficiency and stability of the two algorithms one by one according to the results.

Determine the Number of Video Comparisons

In order to determine the number of times each video is compared in this algorithm, we do the following data simulation: In the interval $[\min, \max]$, we randomly select *sum* different numbers as the actual abnormality level *score* of each video, Then we follow the flow chart 1 to simulate the label, The results of each marked in accordance with the provisions of zero-error mode: According to the actual grade points directly determine the comparison results, the rules are as follows:

$$result_{A,B} = \begin{cases} 1 & score_A > score_B \\ 0.5 & score_A = score_B \\ 0 & score_A < score_B \end{cases} \tag{6}$$

We can think that the number of times in the zero error modes is actually the optimal number of times in the actual operation.

After each round of comparison, these analog videos are ranked according to the ELO rating and the actual rating, and then we compute the Spearman correlation (Referred to as SROCC coefficient which is mainly used to solve the ranking data and sequential data-related issues) of the two ranking results. SROCC correlation coefficient is calculated as follows:

$$\rho_s = 1 - \frac{6 \sum d_i^2}{n(\text{n}^2 - 1)} \tag{7}$$

In Eq.7, d_i is the ranking gap between the two algorithms, n is the number of samples. From Eq.7, the higher the correlation of the two algorithms is, the closer the linear system is. The detailed simulation results are listed in Table 2:

Table 2   SROCC coefficients for simulation experiments

Number of Sample / Number of Comparison	500	1000	2000
5	0.9496	0.9511	0.9505
10	0.9780	0.9765	0.9763
15	0.9829	0.9829	0.9821
20	0.9870	0.9819	0.9831

It can be seen from the table above that when the total number of samples is 2000, the ranking result and the actual ranking result have reached 0.98 or more after each video is marked 20 times.

Algorithm Efficiency Analysis

If the total number of video samples is M , the number of viewers is n , In the MOS score method, each video is labeled m_1 times, the time required for each mark is t_1 . In this algorithm, each observer selects a pair of videos to mark each time, and each video is labeled with m_2 times, and the time required for each labeling is t_2 .The efficiency of the MOS algorithm is T_1 which can be expressed as:

$$T_1 = (M \times m_1) \times t_1 \tag{8}$$

The efficiency of the ELO algorithm is T_2 which can be expressed as:

$$T_2 = \frac{(M/2) \times m_2}{N} \times t_2 \tag{9}$$

The MOS score method requires multiple scoring for each video in the video sample, and the more detailed the rating, the longer the viewer will score the video each time. But ELO algorithm is based on Bradley-Terry model annotation method and the observer only needs to mark the abnormal grade level of two video at a time, so it will greatly reduce the observer's thinking time to speed up labeling efficiency. In the actual operation process, $t_1 \in [6,10]$, $t_2 \in [3,7]$, So we can set $t_1 = 10s$, $t_2 = 5s$ .In the MOS method, each video should be labeled twice, so we can set $m_1 = 2$ .Refer with Table 2 we can set $m_2 = 20$ .The experiment invoked 10 observers, so $n = 10$ .The efficiency of this algorithm and MOS scoring method are listed in Table 3:

Table 3   Comparison of algorithm efficiency

Number of videos	MOS[s]	ELO[s]
500	10000	2500
1000	20000	5000
2000	40000	10000

From the table data we can see that with the increase in the number of video time difference between the two algorithms are also growing. Therefore, with the increase of the order of magnitude of the video, the advantage of this algorithm over the traditional algorithm in terms of time consumption is also increasing.

Stability Analysis

According to reference [10], we can see that the stability of the ranking algorithm can be regarded as the difference between the video with the exception level and the video with the exception level lower

than the middle video, distinguish between the higher the distance between the more scattered to avoid frequent fluctuations in ranking. For a set of video samples with a total of M , the discriminant D is calculated as follows:

$$D = \frac{\sum_{i=1}^{M-1} \sqrt{(V_{i+1} - V_i)^2 + 1}}{\sqrt{2M^2 - 2M - 1}} \tag{10}$$

In Eq.10: $V_i = M \times (1 - \frac{|V_i' - V_1'|}{V_1' - V_M'})$, V' is the original index value for the sample.

The results calculated by the Eq.10 are listed in Table 3, it can be seen that this algorithm has better stability than MOS method.
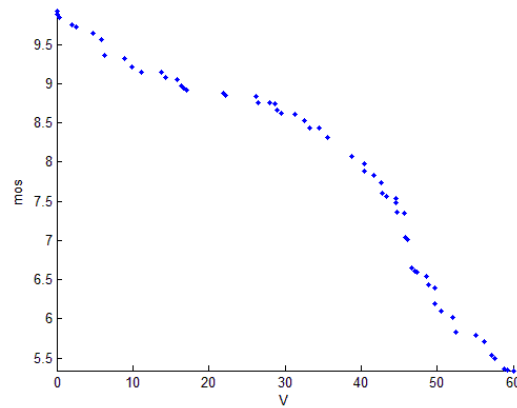
Table 4   Distinction

Statistics	ELO	MOS
maximum	70.89	9.78
minimum	41.23	5.07
average	26.53	7.68
distinction	1.204	1.077

Fig.2 (a) and Fig.2 (b) show the scatter-grams of the ranking scores of the two algorithms respectively. The abscissa is the normalized value V for each video exception level score. The figure show that the results of this algorithm in the abnormal high-grade video score difference is relatively large, and the proportion is relatively small. It is best to distinguish between videos with lower exception levels. Figure b shows the MOS score distribution is more uniform, the higher the abnormal level of video is not a good distinction. So this algorithm is more stable.
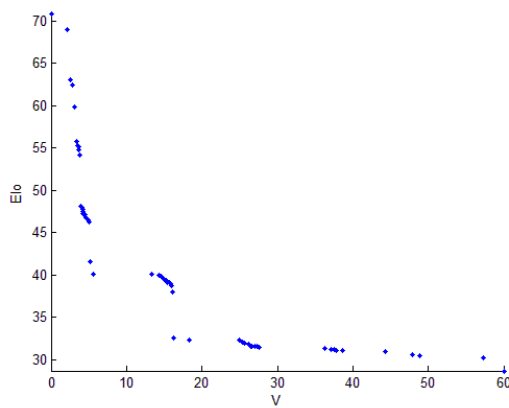
(a) ELO Top 60 Video Scatter Plots   (b) MOS Top 60 Video Scatter Plots

Figure 2.   Ranking of the scatter chart

Conclusion

This paper presents a mass video quality annotation and ranking method based on Bradley-Terry mode, the main research significance is as follows:1) The method of sorting method, each observer only need to mark the abnormal level of the two video without the need for each video are two pairs of comparison, so it will be more efficient.2) Solves the problem that the traditional labeling method cannot fine-sort the massive video quality because of its low classification.3) The experimental results show that the proposed algorithm is more stable and accurate than the traditional MOS scoring method,

and the algorithm running time is more time-saving. The video quality sorting algorithm based on partial order idea is an effective method to deal with the massive video quality sorting, and has certain feasibility, which is suitable for the environment of video big data.

Acknowledgements

References

[1] Y.H.Tian, T.J.Huang, W.Gao: Strategy and Challenge of Intelligent City Video Big Data; China Railway, 2016, No.1, p.82

[2] Z.M.Guo, A.Y.Zhou: Research on Data Quality and Data Cleaning: a Survey; JOURNAL OF SOFTWARE, Vol.13 (2002), No.11, p2076

[3] ITU-R Recommendation. BT.500-13. Methodology for the subjective assessment of the quality of television pictures[S]. http://www.itu.int/rec/R-REC-BT.500-13-201201-I

[4] ITU-T Recommendation P.910.Subjective video quality assessment methods for multimedia applications[S].https://www.itu.int/rec/T-REC-P.910/en.

[5] X.K.Du, X.H.Jiang, C.H.Wang: Subjective Quality Assessment Method of 3DTV Videos; Video Engineering, Vol.37 (2013), No.23, p.241

[6] W.Wang, J.Allebach, Y.Guo. Image quality evaluation using image quality ruler and graphical model [C]// International Conference on Image Processing. QUEBEC CITY, CANADA. 2015.2256-2259

[7] M.Cattelan, C.Varin, D.Firth: Dynamic Bradley-Terry modeling of sports tournaments; Journal of the Royal Statistical Society, Vol.62 (2013), No.1, p.135

[8] C.O.Hu, Z.Q.Liu: Research of Paired-Comparrion Model in News Sorting; Computer Engineering & Software, Vol.32 (2011), No.10, p.24

[9] M.E.Clickman: A Comprehensive Guide to Chess Ratings; Amer.ChessJournal, Vol.3 (1995), p.59

[10] L.P.Yu, Y.T.Pan, Y.H.Wu: Two New Indicators to Compare Different Evaluation Methods' Effect——Based on Times Higher-QS World University Rankings; JOURNAL OF NANJING NORMAL UNIVERSITY (NATURAL SCIENCE EDITION), Vol.31 (2008), No.3, p.135