

## Research on Hand Discrimination for Doudizhu Game

Yipeng Zhang<sup>1,2,a</sup>, Zipeng Chen<sup>1,2,b</sup>, Lanzhou Zheng<sup>2,3,c</sup>, Zhangfei Zhang<sup>2,3,d</sup>,  
Meng Ding<sup>1,2,e</sup>, Kun Meng<sup>1,2,f</sup> and Shuqin Li<sup>1,2,g,\*</sup>

<sup>1</sup>School of Computer, Beijing Information & Science and Technology University, China

<sup>2</sup>Sensing & Computational Intelligence Joint Lab, Beijing Information Science & Technology University, China

<sup>3</sup>Weizhiyu (Beijing) Science and Technology Company Ltd, China

<sup>a</sup>dmm@bistu.edu.cn, <sup>b</sup>bipedalbit@gmail.com, <sup>c</sup>chenzipeng\_1105@163.com, <sup>d</sup>zhenglz@ourgame.com, <sup>e</sup>zzf@smzy.cc, <sup>f</sup>mengkurt@bistu.edu.cn, <sup>g</sup>lishuqin\_de@126.com, \*corresponding author

**Keywords:** Coequal hand quality; Mean value; Standard deviation; Affility propagation algorithm; Inter-cluster average Kullback-Leibler divergence.

**Abstract.** The concept of coequal hand quality is proposed in this paper to judge the fairness while dealing cards in compound Doudizhu competitions, along with two coequal quality hand classifying solutions. The first one is to classify the hands according to mean values and standard deviation among scores of landlord-role players in gaming data, while the second one focuses on dimensionality reduced landlord-role player score probability distribution and a staged clustering algorithm. Experiment confirmed, the dimensionality reduced probability distribution based staged clustering classifying solution performs much better than the statistical solution.

### Introduction

Doudizhu is one of popular Chinese 3-player poker game and the China National Physical Culture Administration has started the China Competitive Poker Championship in September 3, 2016. The championship introduced compound competition mechanism which letting players play game with the same hand and decide the stronger player according to their scores in series of games. [1]

When hands can confirm the fairness of compound competitions between players and share the same properties, those hand are called coequal quality hands. Coequal quality hand classifying solution is designed as the condition of non simultaneous competition and to generate coequal quality hand for different players, increasing gaming interest and the richness of competition mechanism.

The landlord-role player score probability distribution could be the direct metric for individual hand and sufficient real gaming data which contains hands and corresponding landlord-role player scores is in need for coequal quality hand classifying solution.

We got plenty of gaming data from a famous chess and poker game platform. “[T, 3, 5, 5, 5, 7, 7, 8, 8, 8, 8, 9, J, K, K, K, X] [T, T, T, 2, 3, 4, 5, 6, 6, 6, 9, A, D, J, K, Q, Q] [2, 2, 2, 4, 4, 4, 6, 7, 7, 9, 9, A, A, J, J, Q, Q] [3, 3, A] -2” is one of gaming record in gaming data which contains hands of three player including landlord-role player and two farmer-role players as well as three public cards and the score of landlord-role player in the game. T means the 10 card, D and X means the stronger and the weaker trump card respectively. Some studies on coequal quality hand classifying solutions are carried on in this paper according to massive of hand-score data.

### Statistical Coequal Quality Hand Classifying

**Classifying Basis.** Mean value and standard deviation are two of the most fundamental and general demographic characters. Mean value can represent the expectation of a set of data sample while standard deviation is proportional to the volatility. Therefore, these two characters can form one group of the most concise Doudizhu coequal quality hand classifying basis.

**Classifying Solution Designing.** Doudizhu hands are divided according to mean value of landlord-role player scores. We make 5 score mean value sections in total, and to each section there

correspond to a hand classification. Further more, each of these classifications is divided into one great volatility classification and one small volatility classification. 10 classifications with corresponding mean value sections and standard deviations are listed in Table 1.

Table 1. Statistical classifications.

Mean value section	Description	Standard deviation section	Description
(-60.5, -23.5]	landlord great inferiority	(4.0, 35.0]	small volatility
		(35.0, 90.0]	great volatility
(-23.5, -6.5]	landlord small inferiority	(2.0, 15.0]	small volatility
		(15.0, 45.0]	great volatility
(-6.5, 6.5]	balance	(1.0, 7.0]	small volatility
		(7.0, 21.0]	great volatility
(6.5, 23.5]	landlord small superiority	(2.0, 15.0]	small volatility
		(15.0, 45.0]	great volatility
(23.5, 60.5]	landlord great superiority	(4.0, 35.0]	small volatility
		(35.0, 90.0]	great volatility

As what has shown in Table 1, extreme samples which reduce the discrimination of the hands are dropped so that the fairness of compound Doudizhu competition could be ensured.

**Classifying Quality.** Gaming data for solution training are collected before December 20, 2016 from the Doudizhu game platform, which contains approximately 1.8 million gaming records which belong to 0.32 million hand patterns. 37000 hand patterns left after filtering those one played less than 10 times, and only 12000 hand patterns are proper to be used in compound Doudizhu competitions.

More data are collected during December 20, 2016 and February 20, 2017. About 75% of these 0.28 million gaming records varified the statistical solution which confirms the good performance of the statistical coequal quality hand classifying solution.

### Staged Clustering Coequal Quality Hand Classifying

Probability distribution describes the sample set more natively and vividly than mean value and standard deviation. Probability distribution of landlord-role player scores with each hand pattern is chosen to measure the hand quality.

**Data Observation.** Game frequencies of all discrete landlord-role player scores are gained from real gaming data and are shown as Fig. 1. The minimum score and maximum score among 95 discrete scores are -12288 and 6144 respectively. The frequency distribution is approximate to a Gaussian distribution with the score 12 as the symmetry axis.

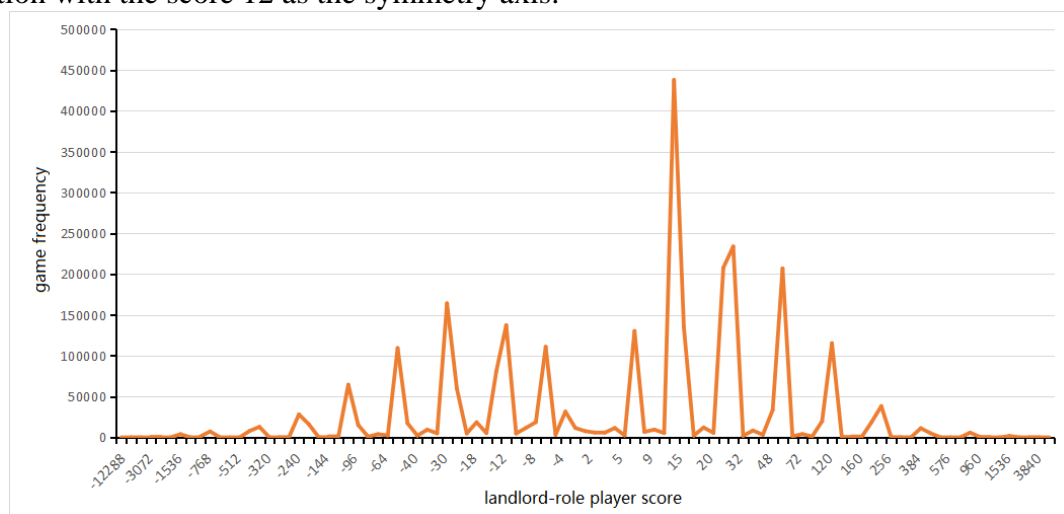


Fig. 1. Frequency distribution of landlord-role player scores.

Discrete scores are merged into several sections to reduce the dimension number of (hand, quality) sample features. Feature extraction solutions are proposed with sample number averaged among score sections which correspond to feature dimensions for clustering algorithm.

**Feature extraction.** Since the frequency distribution of landlord-role player scores contains some peaks, peak scores should generally divided into limited number of different score sections. 3 artificial score section division solutions namely dimension reductive feature extraction solutions are gained. 7-dimension, 10-dimension and 13-dimension feature extraction solutions are shown in Table 2.

Table 2. Feature extraction solutions.

dimension number	7	10	13
section 1	[-12288, -32)	[-12288,-59)	[-12288,-96)
section 2	[-32, -12)	[-59,-17)	[-96,-32)
section 3	[-12, 10)	[-17,-6)	[-32,-30)
section 4	[10,12)	[-6,6)	[-30,-15)
section 5	[12,24)	[6,7)	[-15,-8)
section 6	[24,48)	[7,13)	[-8,-2)
section 7	[48,6144]	[13,25)	[-2,10)
section 8		[25,49)	[10,12)
section 9		[49,120)	[12,20)
section 10		[120,6144]	[20,24)
section 11			[24,40)
section 12			[40,60)
section 13			[60,6144]

**Classify Quality.** (hand, quality) samples are preliminary clustered by affinity propagation (AP) algorithm[2] and the output clusters are further merged by iterative hierarchical clustering algorithm[3][4]. Since Kullback-Leibler divergence[5] is one of the most popular metric of probability distribution similarity, the average K-L divergence between (hand, quality) samples and the cluster centers is chosen to measure the cohesion of clusters. Further more, the inter-cluster average K-L divergence which called ckl in this paper is used to measure the quality of clustering output and it is describe as the following formula.

$$ckl = \frac{\sum_{k=1}^m \left[ \left( \sum_{i=1}^{n_k} D(s_{i,k} \| c_k) \right) / n_k \right]}{m}$$

$s_{i,k}$  means the probability distribution of the  $i$ th sample in the  $k$ th cluster,  $c_k$  means the probability distribution of the  $k$ th cluster center,  $n_k$  means the sum of samples in the  $k$ th cluster,  $m$  means the sum of output clusters.

Perform of our 3 feature extraction solutions with AP algorithm are measured by ckl and the result is shown in Table 3.

Table 3. Clustering results and arguments of feature extraction solutions.

Dimension number	cluster number	ckl	AP.Preference
7	5613	-	default
<b>10</b>	<b>1728</b>	<b>0.124519</b>	<b>default</b>
13	1279	0.2484	default

The 7-dimension feature solution outputs too many clusters which recude the discrimination degree between clusters. The 10-dimension solution and the 13-dimension solution share the same scale of cluster number while the 10-dimension solution has apparent less ckl which means that the 10-dimension solution outputs clusters with more great cohesion.

Finally, we choose the 10-dimension feature extraction solution and the -0.8 preference argument for AP algorithm which explicit the dimensionality reduced probability distribution based staged clustering classifying solution.

## Contrast Experiments and Results

Mid-value of score mean value and standard deviation section are chosen as theory cluster centers, so that the statistical coequal quality hand classifying solution can be compared with the probability distribution based coequal quality hand classifying solution. Besides, the cluster centers are revised by searching closest hand sample and replacing the cluster centers with these hand samples. Contrast experiment is carried on, setting the cluster number of the probability distribution based coequal quality hand staged clustering classifying algorithm to be 10 and the result is shown in Table 4.

Table 4. Classifying solution contrast.

Classifying solution	ckl
Statistical solution	1.871678
Staged clustering solution	0.5

Obviously, the staged clustering solution is of great advantage in Doudizhu coequal quality hand classifying scenario.

## Conclusion

Concept of coequal hand quality is explicated in this paper for solving fairness problem in compound Doudizhu competitions. Landlord-role player score probability distribution is processed in this paper to taking the game sum proportions of different score sections as hand features and 10-dimension feature extraction solution is chosen to be the final solution for the probability distribution based coequal quality hand staged classifying algorithm. And the experiment result indicates the apparent advantage of the probability distribution based staged clustering coequal quality hand classifying solution than the statistical coequal quality hand classifying solution. Our works lay foundation of coequal quality hand classifier implementation in the future, and are of great potentiality for other poker games.

## Acknowledgements

In this paper, the research was sponsored by Standard Project of Science Plan of Beijing Municipal Education Committee (KM201611232014), National Natural Science Foundation of China (No.61502039), the special bidding project of teaching & education reform (2017JGZB08) and 2017 Beijing Information Science and Technology University Graduate Student Science and Technology Innovation Project (No.5121723407).

## References

- [1] Writing group of Chinese Two-on-one Poker Competition Rules. Chinese Two-on-one Poker Competition Rules [M]. People's Physical Press, 2011.
- [2] Frey B J, Dueck D. Clustering by passing messages between data points[J]. science, 2007, 315(5814): 972-976.
- [3] Johnson S C. Hierarchical clustering schemes[J]. Psychometrika, 1967, 32(3): 241-254.
- [4] Jigui Sun, Jie Liu, Lianyu Zhao. Clustering Algorithms Research [J]. Journal of Software, 2008, 19(1): 48-61.
- [5] Gibbs A L, Su F E. On Choosing and Bounding Probability Metrics[J]. International Statistical Review, 2002, 70(3):419-435.