The learning behavior difference between supervised online learning and unsupervised online learning for K-12 education

Jiyou JIA[a*]

[a]* *Department of Educational Technology, Graduate School of Education, Peking University, China*
* jjy@pku.edu.cn

Abstract: Supervised online learning is defined as the online learning under the requirement, supervision and guidance by the teacher, just as in the traditional classroom education. Unsupervised online learning is defined as the online learning without the supervision and guidance by the teacher. I analyze the multiple-dimensional data gathered from one web-based interactive mathematics learning platform "Lexue 100 (Happy Learning for 100 Points)" using SQL command scripts, SPSS statistics and WEKA. The findings include that the supervised online learners from one school spent much more time on participating in more quizzing activities than the unsupervised ones, though the time spent on every quiz by the supervised ones is less than that by the supervised ones, and their mean exercise score is almost the same as the unsupervised ones due to the system's drilling mechanism. This study suggests that the school teacher plays one important role to facilitate the sustainability and speed of pupils' online learning, and the pure online learning by the pupils themselves without the support and requirement from their teachers cannot guarantee the sustainability of the online learning.

Keywords: supervised online learning, unsupervised online learning, learning behavior, quiz

1.   Introduction

The online learning with interactive web-based systems has become popular from K-12 to higher education worldwide. For example, Khan Academy is one famous website with lecture videos, interactive drills and other forms of learning materials (https://www.khanacademy.org). In China, the long-history tradition of valuing education drives the parents to look for more education chances like extracurricular activities besides the regular school education system for their children. The popularity of Internet, personal computers, smart phones and other information and communication technology enable the students to use all kinds of web-based learning materials and platforms to make up for deficiencies from the regular classroom education, such as lack of instant communication, feedback and comments from the teachers, as well as individualized learning demand. To meet such demands, a lot of educational institutes and companies have developed numerous web-based learning and teaching systems for K-12 and higher education. Due to the vast student population, some companies claimed to have gained a great amount of student users. However, the following question has not been thoroughly studied but very interesting both to the online learning providers and consumers: How long and how frequently can the users learn with the online environment?

Besides free usage by pupils, those online learning systems are also used by some schools and teachers as the complement to traditional classroom education. Thus the second question of this study is: what is the learning behavior difference of supervised learning and unsupervised learning using the interactive web-based systems?

The two research questions are important not only for research community but also for the learning system provider.

2. Learning Platform

2.1 System Functions

"Lexue 100" (http://www.lexue100.com) is a web-based interactive learning system for school mathematics. A large number of corresponding quizzes are designed for the different versions of mathematics textbooks that are used in different provinces and metropolis. Each quiz is composed of a series of gap-filling or single-choice questions. So the question's standard answer is predefined and can be compared with the user's trial answer. Thus the positive or negative feedback can be instantly given to the user, as soon as he or she submits the trial answers. Answering quizzes is the main learning behavior in this system.

Based on behaviorism and other learning theories, three learning strategies are adopted by this system: Individualized Adaptation, Incremental Mastery and Interactive Discovery. According to the system's website, they are defined as the following: "Individualized adaptation means that every student can select the quizzes according to his or her own demand. Incremental mastery means that in every quiz, the students should correctly answer every composing question in the sequence, otherwise he or she cannot go on to answer next question. Interactive discovery means that the student can ask for help from the online teaching assistant by answering difficult questions."

To interest the students, the system also delivers scores and credits to students, after they completed one quiz. The amount of score and credits is proportional to the question amount, the time spent on completing the quiz, and the proportion of the questions that are correctly answered by the first trial. Any user can be registered to the system for free of charge, take part in quizzes, and get corresponding credits and amount of virtual currency. For three years this learning system has been used by free users. Some middle schools and primary schools are also invited by the company to use the system for free. In the junior middle school C located in south-west China, one mathematics teacher Y has been invited to use this system to facilitate his teaching and the students' learning for free since September 2015. Because he taught two classes in Grade one of this junior middle school, he selected the corresponding quizzes for the textbook, and required the students from one class as the experiment class to do some quizzes before the lecture and after the lecture using their personal computers or tablet computers at home for every week. He checked the students' online learning activities almost every day. Therefore the online learning of the experiment students from the school C is supervised learning, which can be defined as the online learning under the requirement, supervision and guidance by the teacher, just as in the traditional classroom education. On the contrary, unsupervised online learning is defined as the online learning without the supervision and guidance by the school teacher. The other users of this system, who learn by themselves and without the supervision from their school teachers, are unsupervised students. In order to investigate the two questions regarding to web-based learning systems, I analyze the data from the Lexue100 system as an example.

2.2 User Data

In Lexue100 system, all users' learning activities are tracked and stored in the server database with 281 relational tables. The tables are interleaved with each other. To answer the research questions, I first scrutinized the structures of all tables in details, as well as the complicated relations among them. Then I analyzed the data from those tables dated from January 15, 2013 to May 5, 2016. Among those tables, only those related with users' information and quiz activities are investigated, because doing quizzes is the main learning behavior supported by this system. The records of those tables count from thousands to millions. The biggest table recording all users' quiz activity has more than 33 million records.

In our previous work on MOOC (Massive Open Online Course) and the MOOC users' learning behavior (Jia, Miao & Wang, 2014; Jia & Wang. 2015), we summarized the learning behavior of every MOOC user with a newly designed meta table. In this table, the term "duration" is proposed to describe one user's online time span, and is defined as the time difference between the last login time and the first login time. Similarly, in this paper I summarize the learning behavior of every user with a newly designed table "student_description". The composing fields and their

meanings are listed in Table 1. Its field values are set by executing SQL commands script to calculate other existing tables.

Table 1: The composing fields and their mearnings in the table "student_description"

Field name	Field meaning
id	The user's identification number
first_time	The first time to do the quizzes
last_time	The last time to do the quizzes
duration	The time span between the last time and first time = last_time - first_time
quizzes	The number of doing the quizzes
scores_sum	The sum of scores the user received
scores_mean	The average score = scores_sum/counts
credits_sum	The sum of credits the user received
credits_mean	The average credit = credits_sum/counts
usetime_sum	The sum of time of the user's doing quizzes
usetime_mean	The average time of the user's doing the quizzes= usetime_sum/counts
correctness_mean	The mean of the rate the user correctly answer the questions in the quiz
time_bias_mean	The mean of time bias the user do the quizzes. Time bias is the allowed time period for one quiz minus the time spent on writing one quiz. If the result is positive, the user completed the quiz ahead of the allowed time, otherwise negative, after the allowed time. So this field indicates the speed the user completes one quiz compared with the allowed time.

Based on this student description table, the learning behavior of a specific users' group can be further analyzed, for example, the supervised learners and unsupervised learners. The experiment students from School C are regarded as supervised learners, while the others are regarded as unsupervised learners. The group's mean and standard deviation can be calculated by executing SQL commands script. The correlation among indicators in a group can be calculated with SPSS statistics. The clustering of the learners in a group can be calculated with WEKA or other data mining software. In the following sections, I present the statistical description, correlation and clustering result.

3. Statistical description

The learning behaviors of the two groups are described by their mean and standard deviation, as listed in Table 2. The first column is the learner behavior indicators. The second column is the value of unsupervised learners, and the third column is the value of the supervised learners. The fourth column is the times of the third column's value compared with the second column's value.

Table 2 shows that the supervised learners on average get more scores, credits and a higher correctness rate for every quiz than the unsupervised learners, though the advantage is too small to be noted. This finding can be explained by the drilling mechanism of this system, i.e. Incremental Mastery. Every user has to complete the questions in a quiz correctly in order to go to the next quiz, no matter how many times trials are done and how much time is spent on the questions. The same credit and same score are given to the user after he or she completed the quiz. Those three indicators can be classified as precision indicators.

Furthermore, Table 2 shows the surprising difference between supervised learners and unsupervised learners among the sustainability indicators including duration, quizzes number, time, score and credit. The first one is duration. Although the supervised learners just began their learning with the system on September 15, 2015, and the unsupervised learners began on January 5, 2013, the average duration of the supervised learning is 2.6 times much of the unsupervised learning. The second is the quiz number. The supervised learners did 6.2 times quizzes as the unsupervised ones. It is more noticeable that the supervised learners did almost 2 quizzes every day within the learning duration, while the unsupervised ones did less than one quiz every day. The third, fourth and fifth are quiz score, credit, and time, respectively. Because for every quiz the almost same score and credit were given, the total score, credit and time are proportional to the quiz number.

Table 2 also shows the speed difference for every quiz. The supervised learners spent less time on every quiz than unsupervised learners, i.e. 7 minutes versus almost 10 minutes. The time bias per quiz of the supervised learners is also significantly more than unsupervised ones. Those two indicators demonstrate that supervised learners completed the quiz much faster than the unsupervised ones.

Table 2: The comparison of unsupervised learners with supervised learners

Field name		Unsupervised learners	Supervised learners	Times of the third column's value compared with the second column's value
Users' number		84676	49	-
duration per user	Mean	7449370.2 Seconds = 86.22 Days = 2.87 Months	19450505.1 Seconds = 225.12 Days = 7.5 Months	2.6
	Std. dev.	12044822 Seconds =139.41 Days =4.65 Months	2546789 Seconds =29.48 Days =0.98 Months	0.21
quizzes per user	Mean	62.4	389.4	6.2
	Std. dev.	138.8	123.2	0.89
quizzes per user per day within the duration	Mean	0.72	1.73	2.4
	Std. dev.	4.0	0.5	0.12
scores sum per user	Mean	5980.7	37496.7	6.3
	Std. dev.	15926.5	12395.1	0.78
scores per user per quiz	Mean	93.6	95.3	1.0
	Std. dev.	38.6	4.87	0.13
credits sum per user	Mean	273.0	2018.4	7.4
	Std. dev.	1654.1	740.9	0.45
credits per user per quiz	Mean	4.5	4.6	1.0
	Std. dev.	7.7	1.1	0.14
quiz time sum per user	Mean	29129.5 Seconds = 8.09 Hours	162938.4 Seconds = 45.26 Hours	5.6
	Std. dev.	59752.9 Seconds = 16.60 Hours	68085.2 Seconds =18.91 Hours	1.14
quiz time per user per quiz	Mean	579.2 Seconds = 9.65 Minutes	419.9 Seconds = 7.0 Minutes	0.7
	Std. dev.	437.3 Seconds = 7.29 Minutes	123.3 Seconds = 2.05 Minutes	0.28
correctness rate per user per quiz	Mean	89.3%	90.6%	1.0
	Std. dev.	7.2%	4.1%	0.57
time bias per user per quiz	Mean	35.4 Seconds	283.4 Seconds =4.7 Minutes	8.0
	Std. dev.	883.6 Seconds =14.73 Minutes	165.5 Seconds =2.76 Minutes	0.19

After scrutinizing the users, I find 10993 users with duration=0, who did the quiz for only one time but not anymore, and may be classified as trial users. If those 10993 trial users are excluded from the other really unsupervised learners doing more than one quiz, the comparison of really unsupervised learners with supervised learners is shown in Table 3. The data in Table 3 show that all precision, sustainability and speed indicators of really unsupervised learners are improved, but the improvement is still too trial compared with the supervised learners. In other words, the more than 10 thousand trial

users' single probe of this system did not have great impact on other unsupervised users' cumulative performance.

Table 3: The comparison of really unsupervised learners with supervised learners

Field name		Unsupervised learners doing more than one quiz	Supervised learners	Times of the third column's value compared with the second column's value
Users' number		73683	49	-
duration per user	Mean	8560765 Seconds = 99.08 Days = 3.3 Months	19450505.1 Seconds = 225.12 Days = 7.5 Months	2.3
	Std. dev.	12538259 Seconds =145.12 Days =4.84 Months	2546789 Seconds =29.48 Days =0.98 Months	0.21
quizzes per user	Mean	71.5	389.4	5.4
	Std. dev.	146.58	123.2	0.89
quizzes per user per day within the duration	Mean	0.72	1.73	2.4
	Std. dev.	4.3	0.5	0.12
scores sum per user	Mean	6859.4	37496.7	5.5
	Std. dev.	16898.2	12395.1	0.78
scores per user per quiz	Mean	94.0	95.3	1.0
	Std. dev.	41.03	4.87	0.13
credits sum per user	Mean	313.0	2018.4	6.4
	Std. dev.	1769.73	740.9	0.44
credits per user per quiz	Mean	4.5	4.6	
	Std. dev.	8.21	1.1	0.14
quiz time sum per user	Mean	33350.2 Seconds = 9.26 Hours	162938.4 Seconds = 45.26 Hours	4.9
	Std. dev.	62974.50 Second = 17.49 Hours	68085.2 Seconds =18.91 Hours	1.14
quiz time per user per quiz	Mean	540.3 Seconds = 9.00 Minutes	419.9 Seconds = 7.0 Minutes	0.8
	Std. dev.	346.85 Seconds =5.78 Minutes	123.3 Seconds = 2.05 Minutes	0.28
correctness rate per user per quiz	Mean	89.0%	90.6%	1.0
	Std. dev.	6.92%	4.1%	0.57
time bias per user per quiz	Mean	51.9 Seconds	283.4 Seconds =4.7 Minutes	5.5
	Std. dev.	934.39 Seconds = 15 Minutes	165.5 Seconds =2.76 Minutes	0.19

Besides collective performance expressed by the group mean, both Table 2 and 3 show the standard deviation's difference between the two groups. The unsupervised group is much larger than the supervised group, and the standard deviation of all indicators except quiz time sum per user of the former group is much greater than the deviation of the latter group.

4. Correlation

Correlation — Figure 1

		duration	quizzes	scores_sum	scores_mean	credits_sum	credits_mean	usetime_sum	usetime_mean	correctness_mean	time_bias_mean
duration	Pearson Correlation	1	.543**	.530**	.140	.494**	.228	.343*	-.150	.122	.388**
	significance		.000	.000	.339	.000	.116	.016	.302	.404	.006
	N	49	49	49	49	49	49	49	49	49	49
counts	Pearson Correlation	.543**	1	.990**	.353*	.852**	.238	.749**	-.052	.348*	.600**
	significance	.000		.000	.013	.000	.099	.000	.724	.014	.000
	N	49	49	49	49	49	49	49	49	49	49
scores_sum	Pearson Correlation	.530**	.990**	1	.472**	.897**	.325*	.685**	-.131	.451**	.667**
	significance	.000	.000		.001	.000	.023	.000	.368	.001	.000
	N	49	49	49	49	49	49	49	49	49	49
scores_mean	Pearson Correlation	.140	.353*	.472**	1	.662**	.779**	-.111	-.620**	.918**	.707**
	significance	.339	.013	.001		.000	.000	.448	.000	.000	.000
	N	49	49	49	49	49	49	49	49	49	49
credits_sum	Pearson Correlation	.494**	.852**	.897**	.662**	1	.676**	.471**	-.275	.616**	.832**
	significance	.000	.000	.000	.000		.000	.001	.055	.000	.000
	N	49	49	49	49	49	49	49	49	49	49
credits_mean	Pearson Correlation	.228	.238	.325*	.779**	.676**	1	-.109	-.457**	.767**	.680**
	significance	.116	.099	.023	.000	.000		.454	.001	.000	.000
	N	49	49	49	49	49	49	49	49	49	49
usetime_sum	Pearson Correlation	.343*	.749**	.685**	-.111	.471**	-.109	1	.572**	.012	.124
	significance	.016	.000	.000	.448	.001	.454		.000	.936	.395
	N	49	49	49	49	49	49	49	49	49	49
usetime_mean	Pearson Correlation	-.150	-.052	-.131	-.620**	-.275	-.457**	.572**	1	-.396**	-.542**
	significance	.302	.724	.368	.000	.055	.001	.000		.005	.000
	N	49	49	49	49	49	49	49	49	49	49
correctness_mean	Pearson Correlation	.122	.348*	.451**	.918**	.616**	.767**	.012	-.396**	1	.529**
	significance	.404	.014	.001	.000	.000	.000	.936	.005		.000
	N	49	49	49	49	49	49	49	49	49	49
time_bias_mean	Pearson Correlation	.388**	.600**	.667**	.707**	.832**	.680**	.124	-.542**	.529**	1
	significance	.006	.000	.000	.000	.000	.000	.395	.000	.000	
	N	49	49	49	49	49	49	49	49	49	49

**. significant at .01 level (two tailors)
*. significant at .05 level (two tailors)

Figure 1. The bi-variance correlation for supervised learners

Correlation — Figure 2

		duration	counts	scores_sum	scores_mean	credits_sum	credits_mean	usetime_sum	usetime_mean	correctness_mean	time_bias_mean
duration	Pearson correlation	1	.619**	.519**	.012**	.226**	-.011**	.676**	-.068**	-.261**	.013**
	Significance		.000	.000	.000	.000	.001	.000	.000	.000	.000
	N	84676	84676	84676	84676	84676	84676	84676	84676	84676	84676
counts	Pearson correlation	.619**	1	.873**	.027**	.330**	-.009**	.833**	-.115**	-.243**	.029**
	Significance	.000		.000	.000	.000	.007	.000	.000	.000	.000
	N	84676	84676	84676	84676	84676	84676	84676	84676	84676	84676
scores_sum	Pearson correlation	.519**	.873**	1	.493**	.736**	.459**	.696**	-.106**	-.193**	.027**
	Significance	.000	.000		.000	.000	.000	.000	.000	.000	.000
	N	84676	84676	84676	84676	84676	84676	84676	84676	84676	84676
scores_mean	Pearson correlation	.012**	.027**	.493**	1	.911**	.977**	-.001	-.135**	.105**	.036**
	Significance	.000	.000	.000		.000	.000	.773	.000	.000	.000
	N	84676	84676	84676	84676	84676	84676	84676	84676	84676	84676
credits_sum	Pearson correlation	.226**	.330**	.736**	.911**	1	.900**	.274**	-.048**	-.070**	.014**
	Significance	.000	.000	.000	.000		.000	.000	.000	.000	.000
	N	84676	84676	84676	84676	84676	84676	84676	84676	84676	84676
credits_mean	Pearson correlation	-.011**	-.009**	.459**	.977**	.900**	1	-.020**	-.048**	.107**	.019**
	Significance	.001	.007	.000	.000	.000		.000	.000	.000	.000
	N	84676	84676	84676	84676	84676	84676	84676	84676	84676	84676
usetime_sum	Pearson correlation	.676**	.833**	.696**	-.001	.274**	-.020**	1	.006	-.286**	.003
	Significance	.000	.000	.000	.773	.000	.000		.103	.000	.442
	N	84676	84676	84676	84676	84676	84676	84676	84676	84676	84676
usetime_mean	Pearson correlation	-.068**	-.115**	-.106**	-.135**	-.048**	-.048**	.006	1	-.151**	-.182**
	Significance	.000	.000	.000	.000	.000	.000	.103		.000	.000
	N	84676	84676	84676	84676	84676	84676	84676	84676	84676	84676
correctness_mean	Pearson correlation	-.261**	-.243**	-.193**	.105**	-.070**	.107**	-.286**	-.151**	1	.039**
	Significance	.000	.000	.000	.000	.000	.000	.000	.000		.000
	N	84676	84676	84676	84676	84676	84676	84676	84676	84676	84676
time_bias_mean	Pearson correlation	.013**	.029**	.027**	.036**	.014**	.019**	.003	-.182**	.039**	1
	Significance	.000	.000	.000	.000	.000	.000	.442	.000	.000	
	N	84676	84676	84676	84676	84676	84676	84676	84676	84676	84676

**. Signficant at .01 level (two tailors).

Figure 2. The bi-variance correlation for unsupervised learners

Correlation — Figure 3

		duration	quizzes	scores_sum	scores_mean	credits_sum	credits_mean	usetime_sum	usetime_mean	correctness_mean	time_bias_mean
duration	Pearson Correlation	1	.604**	.505**	.005	.218**	-.005	.662**	-.017**	-.272**	.001
	Signigicance		.000	.000	.141	.000	.157	.000	.000	.000	.694
	N	73683	73683	73683	73683	73683	73683	73683	73683	73683	73683
counts	Pearson Correlation	.604**	1	.871**	.023**	.325**	-.005	.828**	-.104**	-.255**	.021**
	Signigicance	.000		.000	.000	.000	.181	.000	.000	.000	.000
	N	73683	73683	73683	73683	73683	73683	73683	73683	73683	73683
scores_sum	Pearson Correlation	.505**	.871**	1	.498**	.736**	.472**	.688**	-.100**	-.201**	.021**
	Signigicance	.000	.000		.000	.000	.000	.000	.000	.000	.000
	N	73683	73683	73683	73683	73683	73683	73683	73683	73683	73683
scores_mean	Pearson Correlation	.005	.023**	.498**	1	.917**	.983**	-.006	-.113**	.077**	.022**
	Signigicance	.141	.000	.000		.000	.000	.096	.000	.000	.000
	N	73683	73683	73683	73683	73683	73683	73683	73683	73683	73683
credits_sum	Pearson Correlation	.218**	.325**	.736**	.917**	1	.912**	.267**	-.046**	-.072**	.011**
	Signigicance	.000	.000	.000	.000		.000	.000	.000	.000	.002
	N	73683	73683	73683	73683	73683	73683	73683	73683	73683	73683
credits_mean	Pearson Correlation	-.005	-.005	.472**	.983**	.912**	1	-.015**	-.028**	.074**	.011**
	Signigicance	.157	.181	.000	.000	.000		.000	.000	.000	.004
	N	73683	73683	73683	73683	73683	73683	73683	73683	73683	73683
usetime_sum	Pearson Correlation	.662**	.828**	.688**	-.006	.267**	-.015**	1	.062**	-.302**	-.006
	Signigicance	.000	.000	.000	.096	.000	.000		.000	.000	.113
	N	73683	73683	73683	73683	73683	73683	73683	73683	73683	73683
usetime_mean	Pearson Correlation	-.017**	-.104**	-.100**	-.113**	-.046**	-.028**	.062**	1	-.132**	-.130**
	Signigicance	.000	.000	.000	.000	.000	.000	.000		.000	.000
	N	73683	73683	73683	73683	73683	73683	73683	73683	73683	73683
correctness_mean	Pearson Correlation	-.272**	-.255**	-.201**	.077**	-.072**	.074**	-.302**	-.132**	1	.030**
	Signigicance	.000	.000	.000	.000	.000	.000	.000	.000		.000
	N	73683	73683	73683	73683	73683	73683	73683	73683	73683	73683
time_bias_mean	Pearson Correlation	.001	.021**	.021**	.022**	.011**	.011**	-.006	-.130**	.030**	1
	Signigicance	.694	.000	.000	.000	.002	.004	.113	.000	.000	
	N	73683	73683	73683	73683	73683	73683	73683	73683	73683	73683

**. Significant at .01 level (two tailors)

Figure 3. The bi-variance correlation for unsupervised learners doing more than one quiz

I use SPSS Statistics (V20) to analyze the correlation between any two indicators. The correlation result for supervised learners, unsupervised learners and unsupervised learners doing more than one quiz is shown in Figure 1, 2 and 3, respectively. Those figures show that in the three groups, the positive correlation between any two indicators of duration, quizzes, quiz score sum, quiz time sum and time bias mean is significant at 0.01 or 0.05 level. The credits mean is significantly positively correlated with correctness mean and time bias mean at 0.01 level, but negatively correlated with use time mean at 0.01 level. The score sum is significantly positively correlated with duration, quizzes, credits sum and mean, use time sum, correctness mean and time bias mean at 0.01 level. Those correlations can be explained by the specification mechanism for scores and credits by the system.

5. Clustering

I use the Weka (V3.8.0) from the University of Waikato, New Zealand (http://www.cs.waikato.ac.nz/ml/index.html), a data mining software, to cluster the users descriptive data. The clustering algorithm is Simple KMeans with the default parameters.

      weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

      The clustering result for supervised learners (N=49) is two clusters. The first cluster contains 13 (27%) learners, while the second one with 36 (73%) learners. All the performance indicators of the instances in second cluster are better or much better than those in first cluster. For example, Figure 4 shows the duration in X axis and quizzes in Y axis, and the blue and red instances are the instances in cluster 1 and 2, respectively. The blue ones are located mostly in the lower-left corner and indicate smaller duration and quiz number value, while the red ones are located mostly higher-right and indicate larger duration and quiz number value.

      The clustering result for unsupervised learners (N=84676) is two clusters. The first cluster contains 69105 (82%) learners, while the second one with 15571 (18%) learners. All the performance indicators of the instances in first cluster are worse or much worse than those in second cluster. For example, Figure 5 shows the duration in X axis and quizzes in Y axis, and the blue and red instances are the instances in cluster 1 and 2, respectively. The blue ones are located mostly in the lower-left corner and indicate smaller duration and quiz number value, while the red ones are located mostly higher-right and indicate larger duration and quiz number value.
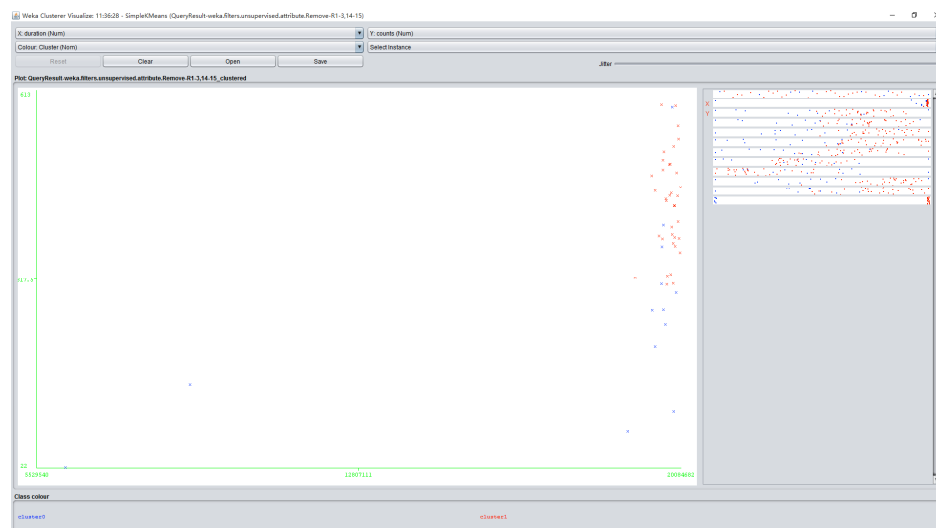

Figure 4. The visualization of clustering result for supervised learners

      The clustering result for unsupervised learners doing more than one quiz (N=73683) is two clusters. The first cluster contains 59040 (80%) learners, while the second one with 14643 (20%) learners. All the performance indicators of the instances in first cluster are worse or much worse than those in second cluster. The clustering result diagram is similar to Figure 2.
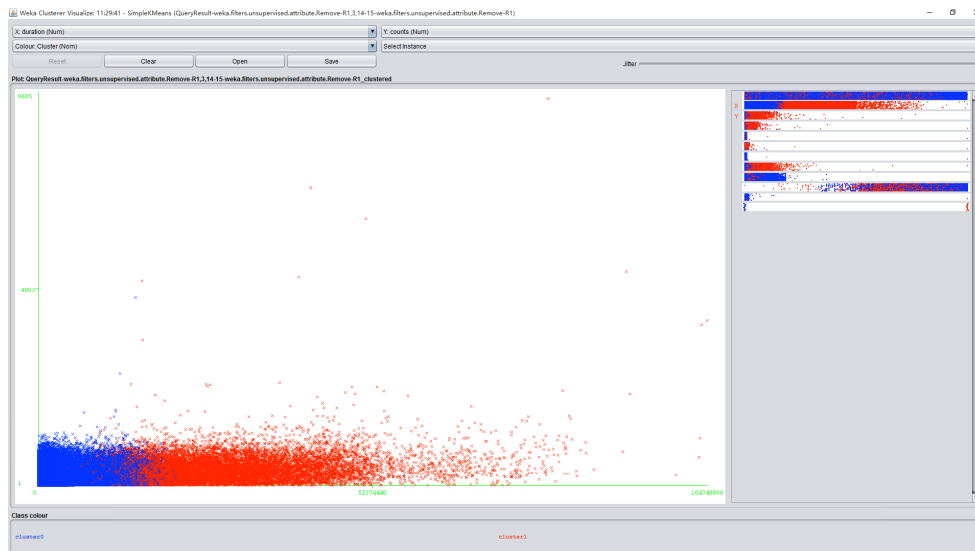
Figure 5. The visualization of clustering result for unsupervised learners

6.   Conclusion and Discussion

Through analyzing the data from Lexue100 system with SQL scripts, SPSS statistics and Weka, I find the supervised online learners from one school spent much more time on participating in more quizzing activities than the unsupervised ones, though the time spent on every quiz of the supervised ones is less than that of the supervised ones, and their mean exercise score is almost the same as the unsupervised ones due to the system's drilling mechanism. The sustainability and speed of the supervised learning are better than that of the unsupervised learning. The reason to explain such findings may be that the school teacher plays one important role to facilitate the sustainability and speed of students' online learning, and the pure online learning by the pupils themselves without the support and requirement from their teachers cannot guarantee the sustainability of the online learning.
        Due to time and literature resource limitation, I have not compared the findings from this specific research with previous related works. Moreover, the learning behavior concerned in this study is only mathematic quizzes for the textbooks. Thus it would be questionable to apply the findings from this study to other web-based online learning platforms. Besides quiz and other learner behavior, interoperability and other important issues should also be considered in learning analytics (Hoel & Chen, 2014).

References
Hoel, T. & Chen, W. (2014). Learning Analytics Interoperability - looking for Low-Hanging Fruits. In: Liu, C.-C. et al. (Eds.): *Proceedings of the 22nd International Conference on Computers in Education*.
Jia, J, Miao, J., & Wang, Q. (2014). The learning behavior and effect analysis of MOOCs based on the big data of six MOOCS from Peking University. *Industry and Information Technology Education (in Chinese)*, 2(9): 23-29.
Jia, J., & Wang, Q. (2015). Using DACP (Data Analyzer of Coursera Platform) to analyze MOOC platform data. In Proceedings of The Fourth International Conference of Educational Innovation through Technology (EITT) 2015 (pp.7-12).