



<b>Title</b>	<b>Cloud-based educational big data application of Apriori algorithm and K-means clustering algorithm based on students' information</b>
<b>Author(s)</b>	<b>Yi, J; Li, S; Wu, M; Au Yeung, HH; Fok, WWT; Wang, Y; Liu, F</b>
<b>Citation</b>	<b>The 4th IEEE International Conference on Big Data and Cloud Computing (BDCloud 2014), Sydney, Australia, 3-5 December 2014. In Conference Proceedings, 2014, p. 151-158</b>
<b>Issued Date</b>	<b>2014</b>
<b>URL</b>	<b><a href="http://hdl.handle.net/10722/209918">http://hdl.handle.net/10722/209918</a></b>
<b>Rights</b>	<b>IEEE International Conference on Big Data and Cloud Computing (BdCloud). Copyright © IEEE Computer Society.</b>

# Cloud-based Educational Big Data Application of Apriori algorithm and K-Means Clustering algorithm based on Students' Information

Jiaqu Yi, Sizhe Li, Maomao Wu, H.H. Au Yeung,  
Wilton W.T Fok

Department of Electrical and Electronic Engineering  
The University of Hong Kong, Hong Kong  
alexyl@eee.hku.hk, wtfok@eee.hku.hk

Ying Wang, Fang Liu

Research institute of Open and Distance education  
The Open University of China  
Beijing 100039, China

**Abstract**—The paper proposes a cloud-based framework to abstract and analyze the meaningful rules among great amount of students' raw information. The authors abstract a set of learning skills based on the course outline from The Open University of China. The authors also present a cloud-based Apriori association algorithm to abstract the rules, followed by a reasonable analysis on educational aspect supported by cloud-based k-means clustering algorithm and learning skills identification.

**Keywords**—cloud; big data; Apriori association algorithm; K-Means clustering; Learning Skills

## I. INTRODUCTION

Cloud Computing and Data Mining are both hot topics nowadays. They bring convenience to diverse fields, including education.

In general, clustering and prediction are two of the most remarkable features of data mining techniques. Unlike traditional analytical methods, data mining could offer more individual-oriented results. The application of data mining techniques in the education field enables numerous possibilities such as comprehensively analyzing the characteristics of each students, predicting success in classes, pinpointing the gifted students and their learning paths [13] etc. In the higher education field, data mining applications have been highly suggested by many researchers such as C Romero and S Ventura in [14] and Luan in [15] to modify or design the curriculum to meet the different needs of students in terms of the learning abilities and knowledge construction.

By these technologies, potentially valuable rules from educational data can be obtained for making decisions and strategies that can optimize the educational resource.

In this project, the authors propose a cloud-based framework to generate the rules. Inside the framework, an Apriori association algorithm is adopted to generate useful rules among the students' grades, followed by reasonable analysis on the generated rules. All the analysis is based on learning skills identification for individual courses and a cloud-based K-means clustering algorithm. The experimental

data come from the Open University of China (OUC). The OUC is a university under the direct administration of the Ministry of Education of the People's Republic of China. It carries out remote education all over China by integrating three networks: a satellite TV network, computer network, and people's network. Nowadays, there are more than 5 million students registered at OUC, and up to 2,679,700 students have graduated from OUC[1]. The research results indicated some significant rules underlying students' school performance and course nature.

## II. LITERATURE REVIEW

### A. Cloud Computing

Cloud computing is a service which can offer the delivery of computing. Resources, software, and information are shared to computers and other devices for achieving high efficiency and economies of scale. It also can be understood as a utility over the Internet[2]. There are three basic service models in cloud platform:

- Infrastructure as a service (IaaS)
- Platform as a service (PaaS)
- Software as a service (SaaS)

Microsoft Windows Azure, Amazon EC2 and Google AppEngine are the well-known cloud platforms used currently. All algorithms in this project are implemented on the Windows Azure cloud platform (renamed as Microsoft Azure after March 25,2014[16]).

- Windows Azure

The Windows Azure Cloud Platform, created by Microsoft, is an application platform for the public cloud. It offers a diverse set of services to application developers. Moreover, Windows Azure can support a large amount of programming languages such as .NET, Java, PHP, Python, Ruby etc. Fig. 1 shows that Azure provides internet accessible application services running in Azure data centers. David Chappell pointed out that windows azure can be utilized by those applications which run in the cloud or local systems[3] as shown in Fig.2.

- Cloud computing for e-Learning

Applying cloud on e-Learning has numerous advantages. Paul POCATILU has shown us the impact on using cloud computing for e-Learning solutions in [17].

Abdullah Alshwaier, Ahmed Youssef and Ahmed Emam theorized in [18] that cloud benefits e-Learning education in terms of cost, efficiency, reliability, security etc. Case studies are given in the paper to prove the idea.

M.Lawanya Shri and Dr. S. Subha implemented an e-Learning application in private cloud in [19], which provided a real case that cloud computing architecture helps achieve scalability, persistent storage, distributed access, efficient resource usage and interoperability of e-Learning system objects[19].

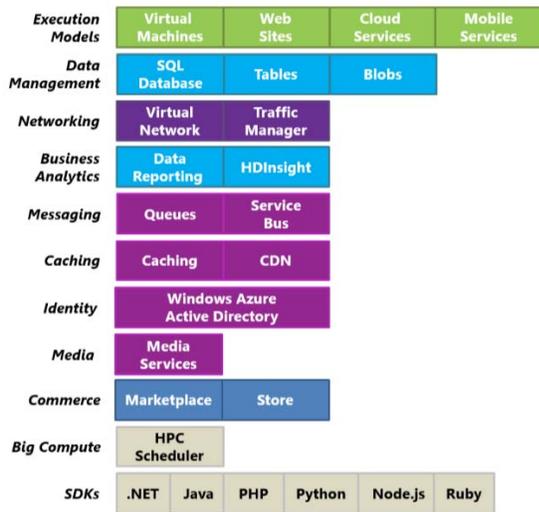


Fig. 1. Azure provides internet accessible application services running in Azure datacenters [20]

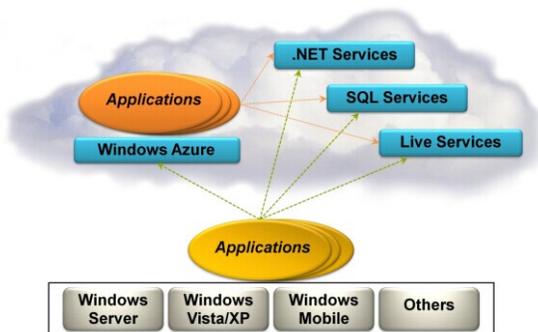


Fig. 2. The Azure Services Platform supports applications running in the cloud and on local systems[3]

## B. Educational Data Mining

Data mining has been used to analyze data in many different fields such as finance, E-commerce etc. It is a crucial part of Knowledge Discovery in Data(KDD) process[6]. In the step of data mining, specific algorithms such as Apriori, k-means and decision-tree can be applied to the data for extracting patterns and potential rules in the data. In education, data mining has brought a new sub-discipline called Educational Data Mining(EDM)[7]. The patterns and rules extracted from EDM systems can be used to make decisions such as course-setting or educational research. Similar to data mining, EDM can be categorized into three groups: classification, clustering and association.

There are many research papers about educational data mining, which are published in authoritative journals and conferences. For instance, Richard A. Huebner present a survey of educational data mining research in [21]. In [21], Huebner describes how data mining can be utilized to analyze the data captured from course management systems. [22] proposes a simple and efficient k-means clustering algorithm which requires a kd-tree as the only major data structure. Ramli, A.A adopts an Apriori algorithm to improve the content of learning portal[22]. Minaei-bidgoli, B Tan, P, Punch, W reveal interesting association rules among the attributes from students and problems in order to optimize online education systems[23]. Merceron, A, Yacef, K. utilize association rules to process learning data and find out whether students use resources to enhance grade and whether their use of such resources affects their grades[24].

## C. Learning Skill on Law

A qualified law graduate should have been developed both legal skills and generic skills after studying at school. A number of researches have listed out certain skills, which are expected from a law degree student. Peden and Riley did a pilot study into employers' perspectives about "law graduates' skills" [25]. After conducting several surveys they believed that the law curriculum should be more focused on "what lawyers need to be able to do" instead of "what lawyers need to know." In their research, skills like "legal skills" (such as legal research and case analysis) are highlighted, as well as other generic skills (such as oral communication and writing).

A study about legal education and professional development named "MacCrate Report" has been done by the "American Bar Association" [12]. It is a report of the task force on Law Schools and profession. This study suggested there are ten kinds of skills:

- Skill 1: Problem Solving
- Skill 2: Legal Analysis and Reasoning
- Skill 3: Legal Research
- Skill 4: Factual Investigation
- Skill 5: Communication
- Skill 6: Counseling
- Skill 7: Negotiation

Skill 8: Litigation and Alternative Dispute-Resolution Procedures

Skill 9: Organization and Management of Legal Work

Skill 10: Recognizing and Resolving Ethical Dilemmas

### III. METHODOLOGY

#### A. K-means Clustering Algorithm

K-means algorithm is one of most important data clustering algorithm. Clustering can be understood as grouping. In this paper, K-means algorithm is used to cluster the courses into different groups according to the required learning skills. K-means algorithm was first raised by James MacQueen and Hugo Steinhaus. Generally, it can be separated into three main parts[9].

1) Centroids Initialization: To initialize the centroids by randomly choosing k observations from dataset.

2) Assignment Step: allocate each observation to the nearest cluster (the distance between its mean and the observation is shortest compared with other clusters)[9].

3) Update Step: recalculate the mean to be the centroid of each new cluster[9].

The pseudo code of the K-means algorithm is displayed as follows:

- 1) Initialize the k centroids
- 2) Calculate the distances between each observation and each centroid.
- 3) Allocate each observation to the nearest centroid.
- 4) Recalculate the mean to be the centroid of each new cluster.
- 5) Repeat 2 to 4 until convergency happens
- 6) Repeat step 1 to step 5 with k from 10 to 2, and pick the best one according to the sum of distance between each observation and centroids.

Fig.3 shows the logic flow of K-means algorithm with a specified K.

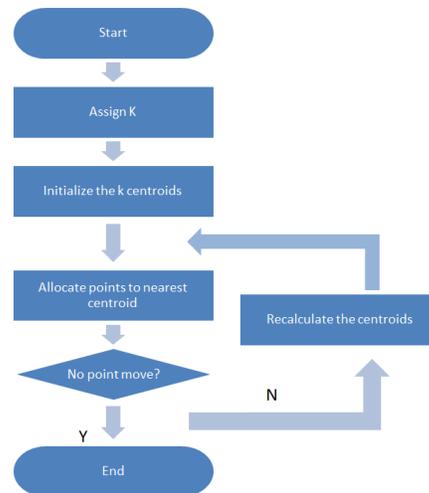


Fig. 3. The logic flow of K-means algorithm with a specified K

#### B. Apriori Association Algorithm[10]

One of the most notable and effective association algorithms is the Apriori algorithm.

##### 1) Related Key Words About Association Rules

a) Item: different values in a transaction. A data set is a collection of transactions[11].

b) Itemset: A set of items is an itemset. Every transaction is an itemset. If an itemset X contains k items, it is called k-itemset. Given a non-empty itemset X, a transaction T in dataset D contains X if  $X \subseteq T$ . Maximum itemset is an itemset which consists of all the items, generally denoted by "I"[11].

c) Association rules: An association rule is an induction rule of the dataset. It can be represented as  $X \Rightarrow Y$ , where  $X \subseteq I$ ,  $Y \subseteq I$ , and X and Y both are not empty,  $X \cap Y = \emptyset$ . Generally, we use following two conditions are used to generate Association rules[11]:

I : Support condition: In transaction dataset D, where X and Y are non-empty, disjoint itemsets, the support condition of transaction  $X \cup Y$  is  $\text{sup}(X \cup Y) / N \geq \rho_s$ ,  $\rho_s$  = support threshold. This formula means the ratio between the amount of transactions  $X \cup Y$  and the total amount N of transactions should be greater than support threshold.

II : Confidence condition: The confidence condition ensures that the mined rules are of high confidence. the rules should satisfy the confidence condition:  $\text{sup}(X \cup Y) / \text{sup}(X) \geq \rho_c$ ,  $\rho_c$  = confidence threshold. It means the probability that itemset Y occurs in a transaction where itemset X occurs, should be more than the confidence threshold.

d) Frequent itemsets: If an itemset satisfies the support condition, it can be called a frequent itemset.

##### 2) Principle of Apriori Algorithm

Apriori algorithm was proposed independently by Agrawal Srikant, Mannila, Toivonen and Verkamo. Apriori algorithm can be divided into two steps:

I : To find out all frequent itemsets. Apriori algorithm utilizes a recursive mode to generate all frequent itemsets.

II: To find out all pairs of itemsets in step 1, which can satisfy the confidence condition. Then, all the association rules can be acquired.

The pseudo code of the above two steps is produced below:

Ck: candidate k-itemset

Lk: frequent k-itemset

STEP 1:

- (1) L1 = find out frequent 1-itemsets in database. And record the amount of each item.
- (2) for (k=2; Lk-1  $\neq \Phi$  ; k++) {
- (3) Ck can be generated by merging pairs of Lk-1.
- (4) for each transaction t in database {
- (5) for each candidate  $c \in C_k$
- (6) if c is in t {c.count++}
- (7) prune those  $c \in C_k$  whose k-1 item-subsets are not all in Lk-1
- (8) Lk = those c in which c.count  $\geq$  min\_support }
- (9) return L = all Lk;

STEP2:

- (1) for (k=1; Lk  $\neq \Phi$  ; k++)
- (2) for (j=2; Lj  $\neq \Phi$  ; j++)
- (3) if x in Lj y in Lk and they satisfy the confidence condition ( $\geq$  min\_conf)
- (4) return  $x \Rightarrow y - x$ .

### C. Learning Skills in Law required by OUC

Based on the curriculum design and course description offered by OUC, the authors identified four basic kinds of skills trained in OUC, namely: Attitudinal skills, cognitive skills, communication skills and relational skills. Table 1 shows the details of the four basic skills. Skills such as Social justice and ethical skills are in the group of "attitudinal skills" while legal analysis and logic & reasoning skills are in the group of cognitive skills. In the group of communication skills, oral presentations and negotiation skills will be assessed. The relational skills group involves skills such as teamwork and self-management. It is very necessary to categorize these skills for further processing the data.

Attitudinal Skills	Cognitive Skills
Ethical Social justice Perspective Creativities	Legal analysis Discipline knowledge Case understanding Logic and reasoning Processing
Communication Skills	Relational Skills
Oral presentations Negotiation Daily communication Legal interview	Approachability Teamwork Self-management

TABLE I. THE DETAILS OF THE FOUR BASIC SKILLS

### D. Research Framework

The proposed research framework is shown in Fig.4. The "Rule Abstraction Module" collects the students' raw information (scores in the courses) from "Students' Profile Database" to generate some raw rules. Meanwhile, the "Clustering Module" is used to group the courses based on the required learning skills. Based on the grouping results and details of skills, we analyze the rules generated by "Rule Abstraction Module" for verification and filtering. Finally, we obtain the useful rules to further instruct education strategies making. The "Rule Abstraction Module" is implemented based on the Apriori algorithm which has been discussed on sub-section B of section II (METHODOLOGY). The "Clustering Module" is designed using K-Means algorithm whose details can be found in sub-section B of section II (METHODOLOGY).

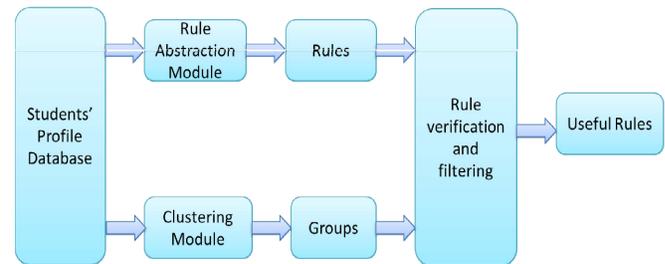


Fig. 4. Research Framework

## IV. EXPERIMENTAL RESULT AND ANALYSIS

### A. EXPERIMENTAL RESULT AND ANALYSIS OF APRIORI ALGORITHM

#### 1) Experimental raw data

The raw data which contains students' grade from OUC is shown in Table II. The data contains student ID, course and scores. All the observed students are from the same major. In total, there are 6688 lines of score information in the raw data. All this data are stored in a CSV document.

Student_id	gender	major	score	courseID	course_name	hometown
711001200037	Female	Law	80	62	Dissertation	Beijing
711001200037	Female	Law	85	372	Legal Logic	Beijing
711001200037	Female	Law	70	395	Real Estate Law	Beijing
711001200037	Female	Law	10	405	Open Education Study Guide	Beijing
711001200037	Female	Law	68	576	Public International Law	Beijing
711001200037	Female	Law	61	585	International Economic Law	Beijing
711001200037	Female	Law	68	599	International Private Law	Beijing
...	...	...	...	...	...	...

TABLE II. THE RAW DATA FROM OUC

### 2) Data pre-processing

In the raw data, some attributes such as gender and hometown are unnecessary, and some records are repetitive and redundant. Hence, the data needs to be cleaned up. And stored into MySQL. MySQL and PHP are used to discretize each score into 5 levels: A(greater than 90), B(< 90 but >= 80),C(< 80 but >=70),D(< 70 but >= 60),F(less than 60).After preprocessing, the clean data generated is shown in Table III.

Student_id	major	score	scoreLV	courseID	course_name
711001200037	Law	80	B	62	Dissertation
711001200037	Law	85	B	372	Legal Logic
711001200037	Law	70	C	395	Real Estate Law
711001200037	Law	100	A	405	Open Education Study Guide
711001200037	Law	68	D	576	Public International Law
711001200037	Law	61	D	585	International Economic Law
711001200037	Law	68	D	599	International Private Law
...	...	...	...	...	...

TABLE III. THE CLEAN DATA

### 3) Data Mining

In the cloud platform, PHP arrays are used to store all candidate 1-itemsets and record the support count of each 1-itemset. The minimum support condition is set to 0.1 which means 10% of the total amount (the total amount is more than 300).The minimum confidence condition is set to 0.7 which is the conditional probability between two frequent itemsets. A part of C1 and L1 generated by Apriori algorithm is displayed in Table IV and Table V respectively. Those candidates whose support counts are lower than the support condition will be pruned.

Itemset	Amount
62:Dissertation (Science of Law) : A	0
62:Dissertation (Science of Law) : B	24
62:Dissertation (Science of Law) : C	72
62:Dissertation (Science of Law) : D	75
62:Dissertation (Science of Law) : F	8
66:Dissertation (Law) : A	0
66:Dissertation (Law) : B	8
66:Dissertation (Law) : C	42
66:Dissertation (Law) : D	38
66:Dissertation (Law) : F	14
.....	.....

TABLE IV. CANDIDATE 1-ITEMSETS C1

Itemset	Amount
62:Dissertation (Science of Law) : C	72
62:Dissertation (Science of Law) : D	75
66:Dissertation (Law) : C	42
66:Dissertation (Law) : D	38
285:Deng Xiaoping's theory and Three Represents Theory: A	37
285:Deng Xiaoping's theory and Three Represents Theory: B	49
364:Jurisprudence: A	41
364:Jurisprudence: B	37
371:Legal Instrument: B	59
371:Legal Instrument: C	87
.....	.....

TABLE V. FREQUENT 1-ITEMSETS L1

After the process of association in Apriori algorithm, 29 association rules are obtained. A part of the result is showed as Table VI.

Rules	Confidence
66:Dissertation (Law) : C =>1708:Science of Criminal Law(Part 1): C	0.73
66:Dissertation (Law) : C =>1712:Administrative Law: C	0.8
66:Dissertation (Law) : C =>50641:Social Practice(Law) : C	0.85
66:Dissertation (Law) : C =>51442:Introduction to Sun Yat-sen's Thoughts: C	0.73
383:Criminology: B =>51442:Introduction to Sun Yat-sen's Thoughts: C	0.81
961:Economic Jurisprudence: B =>1097:Science of Civil Law: A	0.7
961:Economic Jurisprudence: B =>50641:Social Practice(Law) : C	0.72
1097:Science of Civil Law: A =>1712:Administrative Law: C	0.71
1097:Science of Civil Law: A =>51442:Introduction to Sun Yat-sen's Thoughts: C	0.71
1111:Science of Civil Law: A =>1697:Science of Criminal Law(Intemational): B	0.71
1697:Science of Criminal Law(Intemational): B =>382:Base in Law Theory: B	0.76
1711:Science of Criminal Law(Part 2): C =>1712:Administrative Law: C	0.7
.....	.....

TABLE VI. RESULT OF ASSOCIATION RULES

Significant trends can be seen from the results of Apriori algorithm. For example, "961: Economic Jurisprudence: B =>1097: Science of Civil Law: A Confidence=0.70", it can be said that if a student gets B level in Economic Jurisprudence, he will excel in course 1097: Science of Civil Law. Profoundly, it means course 961 may requires some similar skills as course 1097. In another words, students who are able to perform well in both course 961 and 1097, certain parts of his/her learning skills are better than the rest of the students sample, vice versa. Another example, "1711: Science of Criminal Law(Part 2): C =>1712: Administrative Law: C Confidence=0.70", it tells us that a student gets C level in Science of Criminal Law(Part 2), he will probably get a bad level in course 1712: Administrative Law. That is very likely to say, there may be some similarity on the knowledge or skills requirement between course 1711 and course 1712. Students who cannot get a good level in both of the courses, he/she might be lacking of certain knowledge or learning skills, vice versa. According to above analysis, we can realize further

that the result of Apriori algorithm is valuable for educators to make decisions on courses arrangement or adjustment.

## B. EXPERIMENTAL RESULT OF K\_MEANS CLUSTERING ALGORITHM

### 1) Experimental raw data and results

As discussed in sub-section C(Learning Skill in Law) of section II(METHODOLOGY), we identified four basic kinds of skills trained in OUC. We evaluate the courses by this four skills based on the course outline. An example of evaluated results is shown as Fig.5. The scores for the four skills range from 0 to 5. "5" indicates that this ability is highly desired, and "0" means that the corresponding ability is not required by this course. We input this data into "Clustering Module". Inside "Clustering Module" we cluster the courses via K-Means clustering algorithm which has been discussed in sub-section B of section II(METHODOLOGY). Courses which are similar according to the four skills will be grouped together. To demonstrate the results, we utilized "d3.js" to show the clustering result visually as Fig.6. In Fig.6 we can see that courses are divided into 5 groups/clusters by K-Means algorithm. In each cluster, courses are shown as a smallest circle with their course ID on it. The details in cluster 2 are shown in Fig.7.

Among all these clusters, cluster 5 represents the courses which require all high level skills from those four kinds(5;5;5;5). Cluster 1 stands for the courses which require skills with scores pointing to advanced attitudinal skills (5), general cognitive skills (3), low communication and relational skills (1;1). Similar to cluster 1, those courses which require students' skills with score close to advanced attitudinal skills (5), general cognitive skills (3), basic communication skills (2) and low relational skills are mostly found in cluster 2. On the other hand, cluster 3 courses require low attitudinal skills (1), general cognitive and relational skills (3;3) and high communication skills. Cluster 4 represents courses which require advanced attitudinal skills(5), high cognitive skills (4), and general communication and relational skills(3;3).

skill_requirement_id	course_id	attitudinal_skills	cognitive_skills	communication_skills	relational_skills
0	50900	2	5	3	2
1	2051	5	1	1	2
2	372	4	3	2	2
3	2025	2	3	3	3
4	1012	5	3	1	1
5	650	5	2	1	1
6	1267	5	2	1	1
7	576	5	3	2	1
8	2107	5	2	4	3
9	599	5	4	3	3
10	585	5	3	2	1
11	696	5	3	2	1

Fig. 5. Example of evaluated results

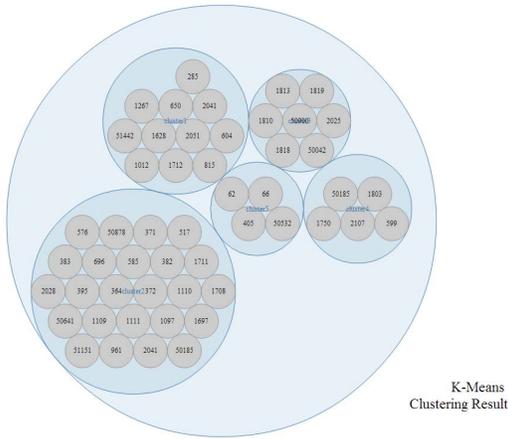


Fig. 6. Result of Clustering Module

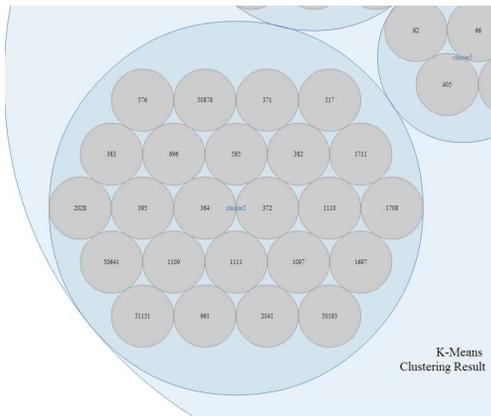


Fig. 7. Details of Cluster 2

In the following section, we will analyze the results of “Rule Abstraction Module” by the results from “Clustering Module”.

### C. RULES VERIFICATION AND FILTERING

#### 1) Cluster 1 and Cluster 5

If a student got a C level in course from cluster 1, he/she was likely to get a C in most of the cluster 5 courses. Courses such as “legal practice”, “dissertation writing” etc. belong to cluster 5. Students who have lower performance in skills like cognitive, communication and relational, even their attitudinal skills are advanced, their performance on academic related courses will be effected directly.

Similar result can be found in the relationship between cluster 2 and cluster 5

#### 2) Cluster 2 and Cluster 5

If a student got a C in most of the cluster 2 courses, he/she would also to get a C in most of the cluster 5 courses.

#### 3) Cluster 2 and Cluster 1

The research team found an interesting phenomenon, when a student got a “A” in courses of cluster 2, he/she was likely to get a “C” in courses from cluster 1. Based on the current data, no explanations could be found to analyze this result. It requires further study with more data.

#### 4) Cluster 1, Cluster 3 and Cluster 4

When a student gets a B from two of any groups, he/she would definitely get a B from the third group courses. cluster 1 has most of the philosophy-related courses while the language courses are found in cluster 3 courses. It is important to note the close relationship between those three clusters. The possible reason is that courses in these three clusters require high cognitive skills, and both cluster 3 and cluster 4 courses require high communication and relational skills.

## V. CONCLUSION

It is tempting to find out the application of big data research in education field. It not only provide an efficient way of analyzing students’ learning skills and academic performance, but more importantly, teachers are able to modify the course content and school work for students based on their school performance. For school principals and education authorities, the result also provides a good reference for designing education curriculum.

The authors present a cloud-based framework to generate the useful rules via K-Means clustering algorithm and Apriori association algorithm. The authors have grouped the learning skills required by graduated law students according to the course outline offered by OUC, namely, attitudinal skills, cognitive skills, communication skills and relational skills. Each groups contains certain relevant learning skills. For instance, skills such as ethical and social justice belongs to the group of attitudinal skills, details please refer to Table 1. This data has been utilized by K-Means clustering algorithm to cluster students into different groups. The clustering results are further used to analyze and filter the rules which are generated by Apriori association algorithm according to the students’ profile. According to the research result, some valuable rules have been found in the current case. For example, the students who have lower performance in skills like cognitive, communication and relational, even their attitudinal skills are advanced, their performance on academic related courses will be effected directly. The current research requires further study to discover more potential and valuable rules to get a clearer picture of the big data application in higher education and more importantly, to further optimize the education resource.

## References

- [1] The Open University Of China  
<http://en.crtvu.edu.cn/index.php/about/33-general-information>
- [2] "The NIST Definition of Cloud Computing". National Institute of Standards and Technology. Retrieved 24 July 2011.
- [3] David Chappell, (October 2008). " Introducing the Azure Services Platform An Early look at Windows Azure, .Net Services, SQL Services, And Live Services ". Chappell & Associates.

- [4] [http://en.wikipedia.org/wiki/Platform\\_as\\_a\\_service](http://en.wikipedia.org/wiki/Platform_as_a_service)
- [5] [http://en.wikipedia.org/wiki/aaS#Infrastructure\\_as\\_a\\_service\\_28aaS.29](http://en.wikipedia.org/wiki/aaS#Infrastructure_as_a_service_28aaS.29)
- [6] Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (1996). "From Data Mining to Knowledge Discovery in Databases". Retrieved 17 December 2008.
- [7] Huebner, Richard A. "A survey of educational data-mining research". *Research in Higher Education Journal*. Retrieved 30 March 2014
- [8] MacQueen, J. B. (1967). "Some Methods for classification and Analysis of Multivariate Observations". *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability I*. University of California Press.
- [9] MacKay, David (2003). "Chapter 20. An Example Inference Task: Clustering". *Information Theory, Inference and Learning Algorithms*. Cambridge University Press
- [10] Rakesh Agrawal and Ramakrishnan Srikant Fast algorithms for mining association rules in large databases. *Proceedings of the 20th International Conference on Very Large Data Bases, VLDB*, pages 487-499, Santiago, Chile, September 1994]
- [11] Agrawal, R.; Imieliński, T.; Swami, A. (1993). "Mining association rules between sets of items in large databases". *Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93*. p. 207.
- [12] American Bar Association Section of Legal Education and Admissions to the Bar, *Legal Education and Professional Development –An Educational Continuum Report of The Task Force on Law Schools and the Profession: Narrowing the Gap*, July, 1992.
- [13] Bael, S., S. H. Hat, and S. C. Parka. "Identifying gifted students and their learning paths using data mining techniques." *Data Mining in E-Learning(Advances in Management Information)* 4 (2006): 191-205.
- [14] C Romero, S Ventura, *Data mining in e-learning*. WIT, 2006.
- [15] Luan, Jing. "Data mining and its applications in higher education." *New directions for institutional research* 2002.113 (2002): 17-36.
- [16] [http://en.wikipedia.org/wiki/Microsoft\\_Azure](http://en.wikipedia.org/wiki/Microsoft_Azure)
- [17] Paul Pocatilu. "Cloud Computing Benefits for E-learning Solutions". *Oeconomics of Knowledge*, 2010, Vol.2(1), p.9
- [18] Abdullah Alshwaier , Ahmed Youssef and Ahmed Emam, "A NEW TREND FOR E-LEARNING IN KSA USING EDUCATIONAL CLOUDS", *Advanced Computing : an International Journal*, 2012, Vol.3(1), p.81
- [18] M.Lawanya Shri, Dr. S.Subha, "An Implementation of e-Learning System in Private Cloud", *International Journal of Engineering and Technology*, 2013, Vol.5(3), p.3036
- [19] <http://azure.microsoft.com/en-us/documentation/articles/fundamentals-introduction-to-azure/>
- [20] Richard A. Huebner, "A survey of educational data-mining research", *Research in Higher Education Journal*, Retrieved 30 March 2014
- [21] Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, Angela Y. Wu, "An Efficient k-Means Clustering Algorithm: Analysis and Implementation", *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, VOL. 24, NO. 7, JULY 2002
- [22] Ramli, A.A. (2005). Web usage mining using apriori algorithm: UUM learning care portal case. In *International conference on knowledge management*, Malaysia, 1-19.
- [23] Minaei-bidgoli, B Tan, P., Punch, W. (2004). Mining interesting contrast rules for a web-based educational system. In *International Conference on Machine Learning Applications*, Los Angeles, USA, 1-8.
- [24] Merceron, A., Yacef, K. (2008). Interestingness Measures for Association Rules in Educational Data. In *International Conference on Educational Data Mining*, Montreal, Canada, 57-66.
- [25] Peden, Elisabeth; Riley, Joellen, "Law Graduates Skills A Pilot Study into Employers Perspectives" [2005] *LegEdRev* 5; (2005) 15(1&2) *Legal Education Review* 87.