World Transactions on Engineering and Technology Education
Vol.5, No.1, 2006

Students' allocation using fuzzy clustering algorithms and Fukuyama and Sugeno's fuzzy cluster validity index

Sani Susanto†, Ernawati‡ & Pandian Vasant*

Parahyangan Catholic University, Bandung, Indonesia†
Atma Jaya University, Yogyakarta, Indonesia‡
University Teknologi Petronas, Perak, Malaysia*

ABSTRACT: Data Structure is a compulsory 4th semester subject offered by the Department of Informatics at Atma Jaya University, Yogyakarta, Indonesia. Consider that there are 121 students enrolled in this subject, all of whom need to be allocated into classes. This fact raises three questions. Firstly, what is the appropriate number of classes? Secondly, how should those students be allocated into classes? Thirdly, which student belongs to which class? The answer to the second and the third questions for a given number of classes was formulated in a previous article by employing the fuzzy clustering concept, one of several tools in data mining. As a development to this prior article, the research presented by the authors in this current article endeavours to answer the first question by employing the Fukuyama and Sugeno's fuzzy cluster validity index.

INTRODUCTION

Consider the following fact: 121 students enrolled in the Department of Informatics at the Atma Jaya University, Yogyakarta, Indonesia, are going to take the subject TF4376 (Data Structure), a compulsory 4th semester unit. Those students must be allocated to, say, n classes. This fact raises three questions. Firstly, what is the appropriate value for n? Secondly, should those 121 students be allocated into classes? Thirdly, which student belongs to which class? In fact, there are common answers to those questions. According to the University's rules, up to 60 students are allowed to be in a class; thus, two classes are needed. Based on students' ID numbers, allocate the first 60 students to are assigned to class A and the remainder 61 students to class B.

However, is there any other more reasonable way to answer those three questions? The answer to the second and third question was given by Susanto, Suharto and Sukapto [6]. This was carried out by applying a data mining technique called *fuzzy clustering* or *fuzzy c-means*. This technique divides several items into several groups (also called *classes* or *clusters*), based on an items' characteristics or attributes.

The research presented in this article is a development of the previous article and tries to answer the first question [6]. The problem formulated by this question belongs to region of *fuzzy cluster validity* problems. Answering it requires a comparison to be carried out between the values of the *fuzzy cluster validity index* for several numbers of the clusters.

This article is organised as follows. The Approach and Methods section describes the approach, methods and concepts applied in order to solve the research questions formulated in the Introduction section. The Results section reports on the information obtained from the application of the approach, methods and the concept employed. In the Discussion section, the authors discuss and interpret the results obtained. The Conclusion section presents the summary of the research results. For the readers' ease and convenience and due to space limitations some tables are placed in the Appendices.

APPROACH AND METHODS IN CLUSTERING

The process of distributing students to classes is called *clustering*, while the class obtained is called a *cluster*. *Clustering* of the 121 students for the subject TF4376 is based on students' mastery level of its prerequisite. In the case of TF4376, its prerequisite subjects are TF2474 (Algorithms and Programming) and TF3276 (Introduction to Data Structure). Students' individual mastery levels of the prerequisites are based on the scores that students achieved. At the Atma Jaya University, students' grades are divided into 11 categories, ie from A (excellent) to E (fail), which corresponds to the score from 4 to 0, as listed in Table 1.

Table 1: Grade and score.

Grade	Score
A	4.00
A^-	3.70
B^+	3.30
B	3.00
B^-	2.70
C^+	2.30
C	2.00
C^-	1.70
D^+	1.30
D	1.00
E	0.00

Scores provide the input information for *clustering*. Successful *clustering* results in *clusters* of students with similar mastery levels of prerequisite subjects. Students' scores of these prerequisites are listed in Table 2 (see Appendices) as *attributes*. The clustering process is based on these attributes. From Table 2, the attribute of the 1st student is represented by the following vector:

$$\mathbf{x}_1 = \begin{pmatrix} 3.00 \\ 2.00 \end{pmatrix} \quad (1)$$

This means that this student attained B and C for TF2474 and TF3276, respectively. The *fuzzy clustering* technique allocates these 121 attributes *vectors*, $x_1,\ldots,x_{121}$, into *c* clusters. This technique is called *fuzzy* clustering, since it gives the *degree of membership* to each cluster for each *attribute vector.* This means that the *fuzzy clustering technique* gives a *suitability level* to each student to belong to each of the *c* clusters. A student who has the highest *degree of membership* to a cluster is assigned to be a member of this cluster. Once *the fuzzy clustering algorithm* receives an input in terms of the attribute vectors, it gives two kinds of vectors as its outputs. The *first* vector, called the *degree of membership vector*, is:

$$U_i = \begin{pmatrix} u_{1i} \\ \vdots \\ u_{ki} \\ \vdots \\ u_{121,i} \end{pmatrix},\ i = 1,2,\ldots, c \quad (2)$$

where U_{ki} represents *the degree of membership* of the k^{th} student to belong to cluster *i*. The *second* vector, called *the cluster centre vector*, is:

$$v_i = \begin{pmatrix} v_{1i} \\ v_{2i} \end{pmatrix},\ i = 1,2,\ldots,c \quad (3)$$

in which v_{ji} represents the (weighted) average of students' scores achieved by students belong to cluster *i* for the j^{th} prerequisite for the subject TF4376. The value of each component in vector v_i represents the student's mastery level of the prerequisites in each cluster.

The component value of vectors U_i and v_i are obtained by solving *the fuzzy clustering problem*, which is basically a constrained optimisation problem in the form as follows:

$$\min J_2(\mathbf{U},\mathbf{v}) = \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ki})^2\left\|\mathbf{x}_k - \mathbf{v}_i\right\|^2 \quad (4)$$

subject to:

$$\sum_{i=1}^{c} u_{ki} = 1, \forall k = 1,2,\ldots,n \quad (5)$$

$$\sum_{k=1}^{n} u_{ki} > 0, \forall i = 1,2,\ldots,c \quad (6)$$

The descriptions of the notation, equation and inequality are as follows:

- The variable *n*, in this case n = 121, represents the number of students taking the subject TF4376;
- The variable *c* represents the number of clusters, of which the value is determined by comparing the values of *fuzzy cluster validity indices* for several number of clusters;
- The matrix $U = (u_{ki})_{nxc}$ consists of *n* rows and *c* columns, of which the elements represent *the degree of membership* of the k^{th} student to undertake the subject TF4376 in the cluster *i*;
- The matrix $v = (v_{ji})_{mxc}$ consists of *m* rows and *c* columns, of which the element represents the (weighted) average of students' score achieved by students belonging to the cluster *i* for the j^{th} prerequisite for the subject TF4376;
- The equation (5) requires that the total degree of membership of each student to belong to the available *c* clusters is 1 or 100%;
- The inequality (6) requires that of all clusters, there must be at least one that a student can belong to.

In extreme conditions, the value of the functional $J_2(U,v)$ in equation (4) is zero, which indicates that the clusters obtained are ideal, since they consist of students with the same mastery level of the prerequisites. The less the value of $J_2(U,v)$, the better that the clustering process is. Some researchers proposed fuzzy clustering algorithms to solve the optimisation problem ([1-5]). For example, the algorithm presented below was proposed by Bezdek [2].

Step 1: Fix c, $2 \le c < \sqrt{n}$; choose any $\xi > 0$ and any norm $\|\ \|$ on R^m; initialise $U^{(0)}$; set *l* = 0.

Step 2: Calculate the c fuzzy cluster centres $\{v_i^{(l)}\}$, where

$$v_{ji}^{(l)} = \frac{\sum_{k=1}^{n}(u_{ki}^{(l)})^2 x_{jk}}{\sum_{k=1}^{n}(u_{ki}^{(l)})^2};\ i=1,\ldots,c;\ j=1,2 \quad (7)$$

Step 3: $l \leftarrow l + 1$; calculate

$$u_{ki}^{(l)} = \begin{cases} \dfrac{1}{\left[\sum_{m=1}^{c}\left(\frac{d_{ki}}{d_{km}}\right)^2\right]}, & \text{if } I_k = \varnothing \\ 0, & \forall i \in I_k^c \text{ if } I_k \neq \varnothing \\ \dfrac{1}{|I_k|}, & \forall i \in I_k \text{ if } I_k \neq \varnothing \end{cases} \quad (8)$$

where $I_k = \left\{i \middle| 1 \le i \le c; d_{ki} = \left\|\mathbf{x}_k - \mathbf{v}_i\right\| = 0\right\}$, $I_k^c = \{1,2,\ldots,c\} - I_k$ and $|I_k|$ is the number of element(s) in I_k.

Step 4: If $\left\|U^{(l+1)} - U^{(l)}\right\| < \xi$ then stop, else go to Step 2.

APPROACH AND METHODS IN CLUSTER VALIDITY

In the case discussed, the fuzzy clustering algorithm can be applied to allocate the 121 students from 2 to n=121 clusters. However, Wu and Yang suggested to cluster those students to 2, …, $\sqrt{n} = \sqrt{121} = 11$ clusters [7]. So what is the most appropriate number of clusters? This question involves finding the optimal number of clusters, which belongs to the *cluster validity* area. In this research, the validity function proposed by Fukuyama and Sugeno, called FS, is employed [7]. This function is defined as follows:

$$FS(c) = \sum_{i=1}^{c}\sum_{k=1}^{n}u_{ki}^2\left\|\mathbf{x}_k - \mathbf{v}_i\right\|^2 + \sum_{i=1}^{c}\sum_{k=1}^{n}u_{ki}^2\left\|\mathbf{v}_i - \bar{\mathbf{v}}\right\|^2 \quad (9)$$

where

$$\overline{\mathbf{v}} = \tfrac{1}{c}\sum_{i=1}^{c}\mathbf{v}_i \qquad (10)$$

The best number of clusters, c*, is found by solving $\max_{2\le c\le\sqrt{n}} \mathrm{FS}(c)$.

RESULTS

Due to space limitations, the application of the fuzzy clustering algorithm (with the help of *MATLAB 6.5* software) is illustrated only for the case of allocating 121 students to c=5 clusters.

The *first* vector, called the *degree of membership vector*, obtained as a result of this algorithm, is U_i (i = 1,2,…,121), as displayed in Table 3 (see Appendices). In this example, the values in the 1st row of this table are interpreted thusly:

$$u_{11} = 0.0017; u_{12} = 0.9927; u_{13} = 0.0009; u_{14} = 0.0015; u_{15} = 0.0033$$

From those values, the 1st student is most appropriate to be in cluster 2, since he/she has the highest degree of membership to this cluster. By the same interpretation, the following students' allocation have been obtained using Table 3:

- The 1st cluster consists of the following student numbers: 6, 10, 17, 18, 21, 22, 24-29, 31-34, 36, 37, 40, 42-44, 46-49, 51-56, 58-62, 72, 73, 82, 83, 85, 87, 91, 94, 98, 103, 105, 108, 110-112, 114, 115 and 118;
- The 2nd cluster consists of student numbers: 1, 2, 39, 41, 45, 65, 71, 88, 95-97, 101, 109, 116 and 120;
- The 3rd cluster consists of student numbers: 4, 7-9, 11, 15, 16, 35 and 100;
- The 4th cluster consists of student numbers: 38, 67, 74, 84, 90, 106, 113 and 121;
- The 5th cluster consists of student numbers: 3, 5, 12-14, 19, 20, 23, 30, 50, 57, 63, 64, 66, 68-70, 75, 77, 79-81, 86, 89, 92, 93, 99, 102, 104, 107, 117 and 119.

The complete allocation for c=1,2,…,$\sqrt{121}$ =11 clusters are listed in Table 4 (see Appendices).

The *second* vector obtained, called *the cluster centre vector,* is v_i (i=1,…,5) as displayed in Table 5.

Table 5: The cluster centre vectors for the five clusters.

v_1	v_2	v_3	v_4	v_5
2.0644	3.0123	1.1603	2.6551	2.0452
1.0140	1.9454	1.9478	3.4230	1.8034

As an example, the interpretations of the values in the 1st column of Table 5 are as follows:

- $v_{11} = 2.0644$, the (weighted) average of the mastery level of the first prerequisite subject (TF2474);
- $v_{21} = 1.0140$, the (weighted) average of the mastery level of the second prerequisite subject (TF 3276).

Calculating the value of the function FS in (9) is shown in Table 6. Since the maximum number of the value of the function FS corresponds to c=5, then the optimal number of clusters is five. Thus, it is suggested that the 121 students enrolled in TF4376 (Data Structure) be clustered into five classes.

DISCUSSION

The fuzzy clustering algorithm can be successfully utilised to allocate the 121 students into clusters. The Fukuyama and Sugeno's *fuzzy cluster validity index* has succeeded in determining the optimal number of clusters. However, the following points must be considered when those two concepts are to be implemented in allocating students into clusters:

- Firstly, based on the results displayed in Table 4 (see Appendices), there is no guarantee that each of the clusters obtained will contain approximately the same number of students for a given number of clusters;
- Secondly, based on the results listed in Table 5, the Fukuyama and Sugeno's index, besides being able to determine the optimal number of clusters, gives options to determine the next optimal number of clusters. This kind of option is beneficial in the case where the optimal number clusters suggested by this index is considered uneconomical.

The correlation coefficient between the values in the first and the second row of Table 5 is 0.3349. This low value indicates that there is no significant relationship between students' scores of the two prerequisites of the subject TF4376 (Data Structure), ie TF2474 (Algorithms and Programming) and TF 3276 (Introduction to Data Structure).

CONCLUSION

If there is a need for the clusters generated by the fuzzy clustering algorithm to contain approximately the same number of students, then this algorithm needs to be modified. Such a modification can be accomplished by performing a minor modification to the stopping rule of Step 4 of Bezdek's algorithm.

ACKNOWLEDGEMENTS

The first author would like to extend his sincere thanks to Dr B.H. Bisowarno (Dean of the Faculty of Industrial Technology), Mr D. Siswanto (Vice Dean for Financial and Administrative Affairs) and Dr B.K. Ane (Head of the Department of Industrial Engineering), all from Parahyangan Catholic University, for their administrative help and moral support in publishing this article.

REFERENCES

1. Al-Sultan, K. and Selim, S.Z., A global algorithm for the fuzzy clustering problem. *Pattern Recognition*, 26, **9**, 1357-1361 (1993).
2. Bezdek, J.C., *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press (1981).
3. Kamel, M.S. and Selim, SZ., New algorithms for solving the fuzzy clustering algorithm. *Pattern Recognition*, 27, **3**, 421-428 (1994).
4. Selim, S.Z. and Al-Sultan, K., A simulated annealing algorithm for the clustering problem. *Pattern Recognition*, 24, **10**, 1003-1008 (1991).
5. Susanto, S., Kennedy, R.D. and Price, J.W.H., A preliminary study of a clustering and assignment problem-based cell formation problem. *Proc. Inter. Conf. on Mfg. Automation*, 1, 95-104 (1997).

6. Susanto, S., Suharto, P. and Sukapto, P., Using the fuzzy clustering algorithm for the allocation of students. *World Trans. on Engng. and Technology Educ.*, 1, **2**, 245-248 (2002).

7. Wu, K-L. and Yang, M-S., A cluster validity index for fuzzy clustering. *Pattern Recognition*, Letters, 26, **9**, 1275-1291 (2005).

APPENDICES

Table 2: Attributes of 121 students based on score achieved for the prerequisite subjects of TF4376 Data Structure.

Student	Score for Programming Algorithm	Score for Introduction to Data Structure		Student	Score for Programming Algorithm	Score for Introduction to Data Structure		Student	Score for Programming Algorithm	Score for Introduction to Data Structure
1	3.00	2.00		42	2.70	1.00		83	3.00	1.00
2	3.00	2.00		43	2.70	1.00		84	3.70	4.00
3	2.00	2.00		44	2.30	1.00		85	2.00	1.00
4	1.00	3.00		45	3.00	2.00		86	2.00	2.30
5	2.00	2.30		46	2.30	1.00		87	2.00	1.00
6	2.00	1.00		47	2.70	1.00		88	2.70	2.00
7	1.00	2.70		48	1.70	1.00		89	2.30	1.70
8	1.30	2.00		49	1.70	1.00		90	2.70	2.70
9	1.30	2.70		50	2.00	2.30		91	1.70	1.00
10	2.00	1.30		51	2.00	1.00		92	2.00	2.00
11	1.00	2.00		52	1.30	1.00		93	1.70	2.00
12	2.30	2.00		53	1.70	1.00		94	2.70	1.00
13	1.70	1.70		54	1.70	1.00		95	3.00	1.70
14	2.30	1.70		55	1.70	1.00		96	3.30	1.70
15	1.00	1.70		56	2.00	1.00		97	2.70	1.70
16	1.00	1.70		57	2.00	1.70		98	2.00	1.00
17	1.30	1.00		58	2.30	1.00		99	2.00	1.70
18	2.00	1.00		59	1.70	1.00		100	1.00	1.70
19	2.00	2.00		60	3.00	1.00		101	3.00	2.70
20	2.00	1.70		61	1.70	1.00		102	2.30	2.00
21	2.30	1.00		62	2.30	1.00		103	1.70	1.00
22	2.30	1.00		63	2.30	1.70		104	1.70	1.70
23	2.30	1.70		64	1.70	1.70		105	1.70	1.00
24	2.30	1.00		65	3.00	2.30		106	2.30	3.30
25	2.30	1.00		66	2.00	1.70		107	2.00	2.00
26	2.30	1.00		67	3.70	3.30		108	2.00	1.00
27	2.00	1.00		68	1.70	1.70		109	3.30	2.30
28	2.30	1.00		69	2.00	2.00		110	1.70	1.00
29	2.00	1.00		70	2.00	1.70		111	2.30	1.00
30	2.00	1.70		71	3.00	1.70		112	1.70	1.00
31	2.30	1.00		72	2.70	1.00		113	2.30	3.70
32	2.30	1.00		73	1.70	1.00		114	2.70	1.00
33	2.00	1.00		74	3.00	3.70		115	2.00	1.00
34	2.30	1.00		75	2.30	1.70		116	3.30	2.00
35	1.00	2.00		76	1.70	1.30		117	2.30	1.70
36	2.30	1.00		77	1.70	1.70		118	2.00	1.00
37	2.00	1.00		78	1.70	1.00		119	2.30	1.70
38	2.00	3.00		79	2.00	2.00		120	3.70	1.00
39	3.00	2.00		80	1.70	1.70		121	1.70	3.30
40	2.30	1.00		81	2.00	1.70				
41	3.00	2.30		82	2.00	1.00				

Table 3: The students' degree of memberships to five clusters.

Student	Degree of Membership to Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
1	0.0017	0.9927	0.0009	0.0015	0.0033
2	0.0017	0.9927	0.0009	0.0015	0.0033
3	0.0361	0.0343	0.0497	0.0143	0.8656
4	0.0979	0.0963	0.4387	0.1703	0.1969
5	0.0827	0.1192	0.1654	0.0811	0.5516
6	0.9878	0.0022	0.0027	0.0007	0.0066
7	0.0841	0.0724	0.5649	0.1024	0.1762
8	0.0134	0.0071	0.9389	0.0054	0.0351
9	0.0847	0.0829	0.4958	0.1230	0.2136
10	0.6712	0.0400	0.0513	0.0117	0.2258
11	0.0128	0.0067	0.9509	0.0057	0.0239
12	0.0704	0.1418	0.0556	0.0336	0.6986
13	0.1273	0.0431	0.2179	0.0198	0.5918
14	0.1059	0.0982	0.0410	0.0180	0.7369
15	0.0464	0.0181	0.8549	0.0130	0.0675
16	0.0464	0.0181	0.8549	0.0130	0.0675
17	0.4251	0.0650	0.2707	0.0322	0.2069
18	0.9878	0.0022	0.0027	0.0007	0.0066
19	0.0361	0.0343	0.0497	0.0143	0.8656
20	0.0253	0.0111	0.0157	0.0035	0.9443
21	0.8675	0.0345	0.0220	0.0081	0.0680
22	0.8675	0.0345	0.0220	0.0081	0.0680
23	0.1059	0.0982	0.0410	0.0180	0.7369
24	0.8675	0.0345	0.0220	0.0081	0.0680
25	0.8675	0.0345	0.0220	0.0081	0.0680
26	0.8675	0.0345	0.0220	0.0081	0.0680
27	0.9878	0.0022	0.0027	0.0007	0.0066
28	0.8675	0.0345	0.0220	0.0081	0.0680
29	0.9878	0.0022	0.0027	0.0007	0.0066
30	0.0253	0.0111	0.0157	0.0035	0.9443
31	0.8675	0.0345	0.0220	0.0081	0.0680
32	0.8675	0.0345	0.0220	0.0081	0.0680
33	0.9878	0.0022	0.0027	0.0007	0.0066
34	0.8675	0.0345	0.0220	0.0081	0.0680
35	0.0128	0.0067	0.9509	0.0057	0.0239
36	0.8675	0.0345	0.0220	0.0081	0.0680
37	0.9878	0.0022	0.0027	0.0007	0.0066
38	0.0701	0.1294	0.1526	0.4549	0.1929
39	0.0017	0.9927	0.0009	0.0015	0.0033
40	0.8675	0.0345	0.0220	0.0081	0.0680
41	0.0387	0.7779	0.0279	0.0710	0.0845
42	0.5060	0.2063	0.0626	0.0348	0.1903
43	0.5060	0.2063	0.0626	0.0348	0.1903
44	0.8675	0.0345	0.0220	0.0081	0.0680
45	0.0017	0.9927	0.0009	0.0015	0.0033
46	0.8675	0.0345	0.0220	0.0081	0.0680
47	0.5060	0.2063	0.0626	0.0348	0.1903
48	0.7374	0.0375	0.0824	0.0145	0.1282
49	0.7374	0.0375	0.0824	0.0145	0.1282
50	0.0827	0.1192	0.1654	0.0811	0.5516
51	0.9878	0.0022	0.0027	0.0007	0.0066
52	0.4251	0.0650	0.2707	0.0322	0.2069
53	0.7374	0.0375	0.0824	0.0145	0.1282
54	0.7374	0.0375	0.0824	0.0145	0.1282
55	0.7374	0.0375	0.0824	0.0145	0.1282
56	0.9878	0.0022	0.0027	0.0007	0.0066
57	0.0253	0.0111	0.0157	0.0035	0.9443
58	0.8675	0.0345	0.0220	0.0081	0.0680
59	0.7374	0.0375	0.0824	0.0145	0.1282
60	0.3458	0.3386	0.0707	0.0505	0.1944

continued ...

Table 3 (continuation).

Student	Degree of Membership to Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
61	0.7374	0.0375	0.0824	0.0145	0.1282
62	0.8675	0.0345	0.0220	0.0081	0.0680
63	0.1059	0.0982	0.0410	0.0180	0.7369
64	0.1273	0.0431	0.2179	0.0198	0.5918
65	0.0387	0.7779	0.0279	0.0710	0.0845
66	0.0253	0.0111	0.0157	0.0035	0.9443
67	0.0709	0.2428	0.0677	0.5061	0.1125
68	0.1273	0.0431	0.2179	0.0198	0.5918
69	0.0361	0.0343	0.0497	0.0143	0.8656
70	0.0253	0.0111	0.0157	0.0035	0.9443
71	0.0391	0.8716	0.0153	0.0170	0.0570
72	0.5060	0.2063	0.0626	0.0348	0.1903
73	0.7374	0.0375	0.0824	0.0145	0.1282
74	0.0208	0.0547	0.0261	0.8610	0.0374
75	0.1059	0.0982	0.0410	0.0180	0.7369
76	0.4956	0.0497	0.1496	0.0196	0.2854
77	0.1273	0.0431	0.2179	0.0198	0.5918
78	0.7374	0.0375	0.0824	0.0145	0.1282
79	0.0361	0.0343	0.0497	0.0143	0.8656
80	0.1273	0.0431	0.2179	0.0198	0.5918
81	0.0253	0.0111	0.0157	0.0035	0.9443
82	0.9878	0.0022	0.0027	0.0007	0.0066
83	0.3458	0.3386	0.0707	0.0505	0.1944
84	0.0703	0.1736	0.0764	0.5719	0.1077
85	0.9878	0.0022	0.0027	0.0007	0.0066
86	0.0827	0.1192	0.1654	0.0811	0.5516
87	0.9878	0.0022	0.0027	0.0007	0.0066
88	0.0529	0.7247	0.0307	0.0359	0.1558
89	0.1059	0.0982	0.0410	0.0180	0.7369
90	0.0633	0.3082	0.0700	0.3917	0.1668
91	0.7374	0.0375	0.0824	0.0145	0.1282
92	0.0361	0.0343	0.0497	0.0143	0.8656
93	0.0782	0.0501	0.2941	0.0294	0.5481
94	0.5060	0.2063	0.0626	0.0348	0.1903
95	0.0391	0.8716	0.0153	0.0170	0.0570
96	0.0580	0.8098	0.0250	0.0342	0.0730
97	0.1088	0.6034	0.0391	0.0320	0.2166
98	0.9878	0.0022	0.0027	0.0007	0.0066
99	0.0253	0.0111	0.0157	0.0035	0.9443
100	0.0464	0.0181	0.8549	0.0130	0.0675
101	0.0609	0.3973	0.0573	0.3526	0.1319
102	0.0704	0.1418	0.0556	0.0336	0.6986
103	0.7374	0.0375	0.0824	0.0145	0.1282
104	0.1273	0.0431	0.2179	0.0198	0.5918
105	0.7374	0.0375	0.0824	0.0145	0.1282
106	0.0224	0.0505	0.0378	0.8379	0.0513
107	0.0361	0.0343	0.0497	0.0143	0.8656
108	0.9878	0.0022	0.0027	0.0007	0.0066
109	0.0486	0.7414	0.0329	0.0922	0.0849
110	0.7374	0.0375	0.0824	0.0145	0.1282
111	0.8675	0.0345	0.0220	0.0081	0.0680
112	0.7374	0.0375	0.0824	0.0145	0.1282
113	0.0235	0.0477	0.0391	0.8430	0.0467
114	0.5060	0.2063	0.0626	0.0348	0.1903
115	0.9878	0.0022	0.0027	0.0007	0.0066
116	0.0301	0.8762	0.0164	0.0308	0.0466
117	0.1059	0.0982	0.0410	0.0180	0.7369
118	0.9878	0.0022	0.0027	0.0007	0.0066
119	0.1059	0.0982	0.0410	0.0180	0.7369
120	0.2224	0.4353	0.0810	0.0855	0.1758
121	0.0764	0.1151	0.1932	0.4416	0.1736

Table 4: The results from clustering 121 students into 2, …,11 clusters.

No. of Clusters	Cluster Number										
	1	2	3	4	5	6	7	8	9	10	11
11	42-43 47 60 72 83 94 114 120	21-22 24-26 28 31-32 34 36 40 44 46 58 62 111	1-2 39 41 45 65 71 88 90 95-96 101 109 116	5 38 50 86 106	67 74 84 113	3 19-20 30 57 66 69-70 79 81 92 99 107	4 7 9 121	8 11 13 15-16 35 64 68 77 80 93 100 104	17 48-49 52-55 59 61 73 76 78 91 103 105 110 112	12 14 23 63 75 89 97 102 117 119	6 10 18 27 29 33 37 51 56 82 85 87 98 108 115 118
10	1 2 39 41 45 65 71 72 73 74 75 88 95 96 109 116 120	67 84 90 101	3 19 69 79 92 107	4-5 7 9 50 86	6 10 17-18 27 29 33 37 48-49 51-56 59 61 73 76 78 82 85 87 91 98 103 105 108 110 112 115 118	20 30 57 66 70 81 99	38 74 106 113 121	12 14 23 63 75 89 97 102 117 119	8 11 13 15 16 35 64 68 77 80 93 100 104	21 22 24-26 28 31-32 34 36 40 42-44 46-47 58 60 62 72 83 94 111 114	
9	7-9 11 15-16 35 100	1 2 39 41 45 65 71 88 90 95-97 101 109 116	6 10 18 21-22 24-29 31-34 36-37 40 44 46 51 56 58 62 82 85 87 98 108 111 115 118	67 74 84	4 38 106 113 121	3 5 12 19 50 69 79 86 92-93 102 107	13-14 20 23 30 57 63-64 66 68 70 75 77 80-81 89 99 104 117 119	42 43 47 60 72 83 94 114 120	17 48-49 52-55 59 61 73 76 78 91 103 105 110 112		
8	12 14 20 23 30 57 63 66 70 75 81 89 97 99 102 117 119	4 38 106 113 121	21-22 24-26 28 31-32 34 36 40 42-44 46-47 58 60 62 72 83 94 111 114	7-9 11 15-16 35 100	1-2 39 41 45 65 71 88 90 95-96 101 109 116 120	67 74 84	3 5 13 19 50 64 68-69 77 79-80 86 92-93 104 107	6 10 17-18 27 29 33 37 48-49 51-56 59 61 73 76 78 82 85 87 91 98 103 105 108 110 112 115 118			
7	67 74 84 106 113	3 5 12 19 38 50 69 79 86 92-93 102 107 121	1-2 39 41 45 65 71 88 90 95-96 109 116 120	13 20 30 57 64 66 68 70 76-77 80-81 99 104	6 10 17-18 21-22 24-29 31-34 36-37 40 42-44 46-49 51-56 58-62 72-73 78 82-83 85 87 91 94 98 103 105 108 110-112 114-115 118	14 23 63 75 89 97 117 119	4 7-9 11 15-16 35 100				
6	12 14 23 60 63 75 83 89 97 102 117 119	2 5 13 19-20 30 50 57 64 66 68-70 77 79-81 86 92-93 99 104 107	1-2 39 41 45 65 71 88 90 95-96 101 109 116 120	38 67 74 84 106 121	4 7-9 11 15 16 35 100	6 10 17-18 21-22 24-29 31-34 36-37 40 42-44 46-49 51-56 58-59 61-62 72-73 76 78 82 85 87 91 94 98 103 105 108 110-112 114-115 118					
5	6 10 17 18 21 22 24-29 31-34 36 37 40 42-44 46-49 51-56 58-62 72-73 82 83 85 87 91 94 98 103 105 108 110-112 114 115 118	1 2 39 41 45 65 71 88 95-97 101 109 116 120	4 7-9 11 15 16 35 100	38 67 74 84 90 106 113 121	3 5 12-14 19 20 23 30 50 57 63 64 66 68-70 75 77 79-81 86 89 92 93 99 102 104 107 117 119						
4	6 10 17-18 27 29 33 37 48-49 51-56 59 61 73 76 78 82 85 87 91 98 103 105 108 110 112 115	21-22 24-26 28 31-32 34 36 40 42-44 46-47 58 60 62 72 83 94 97 111 114 120	3-5 7-9 11-16 19-20 23 30 35 38 50 57 63-64 66 68-70 75 77 79-81 86 89 92-93 99-100 102 104 107 117 119 121	1 2 39 41 45 65 67 71 74 84 88 90 95-96 101 106 109 113 116							
3	1 2 39 41 45 65 67 71 74 84 88 90 95-97 101 106 109 113 116 120	6 10 17 18 21-22 24-29 31-34 36-37 40 42-44 46-49 51-56 58-62 72-73 76 78 82-83 85 87 91 94 98 103 105 108 110-112 114-115 118	3-5 7-9 11-16 19-20 23 30 35 38 50 57 63-64 66 68-70 75-77 79-81 86 89 92-93 99-100 102 104 107 117 119 121								
2	1-5 7-9 12 14 19 23 38 39 41 45 50 63 65 67 69 71 74-75 79 84 86 88-90 92 93 95-97 101-102 106-107 109 113 116-117 119 121	6 10-11 15-18 20-22 24-37 40 42-44 46-49 51-62 64 66 68 70 72-73 76-78 80-83 85 87 91 94 98-100 103-105 108 110-112 114-115 118 120									