

Using the score function of the intuitionistic fuzzy set for the classification problem: "modeling the status of learners' understanding with a subject"

Nguyễn Xuân Thảo
Khoa CNTT-HVNNVN

Bài toán

Có một tập dữ liệu thực về tình trạng hiểu biết của học sinh về môn học Máy điện một chiều, có 5 thuộc tính đầu vào:

Input

STG (The degree of study time for goal object materials),

SCG (The degree of repetition number of user for goal object materials)

STR (The degree of study time of user for related objects with goal object)

LPR (The exam performance of user for related objects with goal object)

PEG (The exam performance of user for goal objects);

Output

UNS (The knowledge level of user) (target value) (4 class)

Tập dữ liệu

- Từ nguồn UCI:
<https://archive.ics.uci.edu/ml/datasets/User+Knowledge+Modeling>
- Tập dữ liệu đã được chia sẵn thành hai tập:
- Tập huấn luyện có 258 đối tượng
- Tập test: 145 đối tượng

Một số công bố liên quan đến tập dữ liệu

Table 14

Comparing the performances of algorithms.

Algorithm	Bayes knowledge classifier	<i>k</i> -NN knowledge classifier			Proposed IKC		
		EU	MA	MI	EU	MA	MI
The average number of misclassified observations	38	26.2	21.7	30.5	3	3	5
The percentages of average error rates (%)	26.2	18.1	15	21	2.1	2.1	3.5
The classification accuracy (%)	73.8	81.9	85	79	97.9	97.9	96.5

Một phần tập dữ liệu train

STG	SCG	STR	LPR	PEG	UNS (Class)
0	0	0	0	0	very_low
0.08	0.08	0.1	0.24	0.9	High
0.06	0.06	0.05	0.25	0.33	Low
0.1	0.1	0.15	0.65	0.3	Middle
0.08	0.08	0.08	0.98	0.24	Low
0.09	0.15	0.4	0.1	0.66	Middle
0.1	0.1	0.43	0.29	0.56	Middle
0.15	0.02	0.34	0.4	0.01	very_low
0.2	0.14	0.35	0.72	0.25	Low
0	0	0.5	0.2	0.85	High
0.18	0.18	0.55	0.3	0.81	High
0.06	0.06	0.51	0.41	0.3	Low
0.1	0.1	0.52	0.78	0.34	Middle
0.1	0.1	0.7	0.15	0.9	High
0.2	0.2	0.7	0.3	0.6	Middle
0.12	0.12	0.75	0.35	0.8	High

Một phần dữ liệu test

STG	SCG	STR	LPR	PEG	UNS
0	0.1	0.5	0.26	0.05	Very Low
0.05	0.05	0.55	0.6	0.14	Low
0.08	0.18	0.63	0.6	0.85	High
0.2	0.2	0.68	0.67	0.85	High
0.22	0.22	0.9	0.3	0.9	High
0.14	0.14	0.7	0.5	0.3	Low
0.16	0.16	0.8	0.5	0.5	Middle
0.12	0.12	0.75	0.68	0.15	Low
0.2	0.2	0.88	0.77	0.8	High
0.16	0.25	0.01	0.1	0.07	Very Low
0.11	0.29	0.2	0.05	0.66	Middle
0.18	0.26	0.05	0.4	0.04	Very Low
0.21	0.32	0.25	0.5	0.8	High
0.13	0.28	0.18	0.75	0.32	Middle

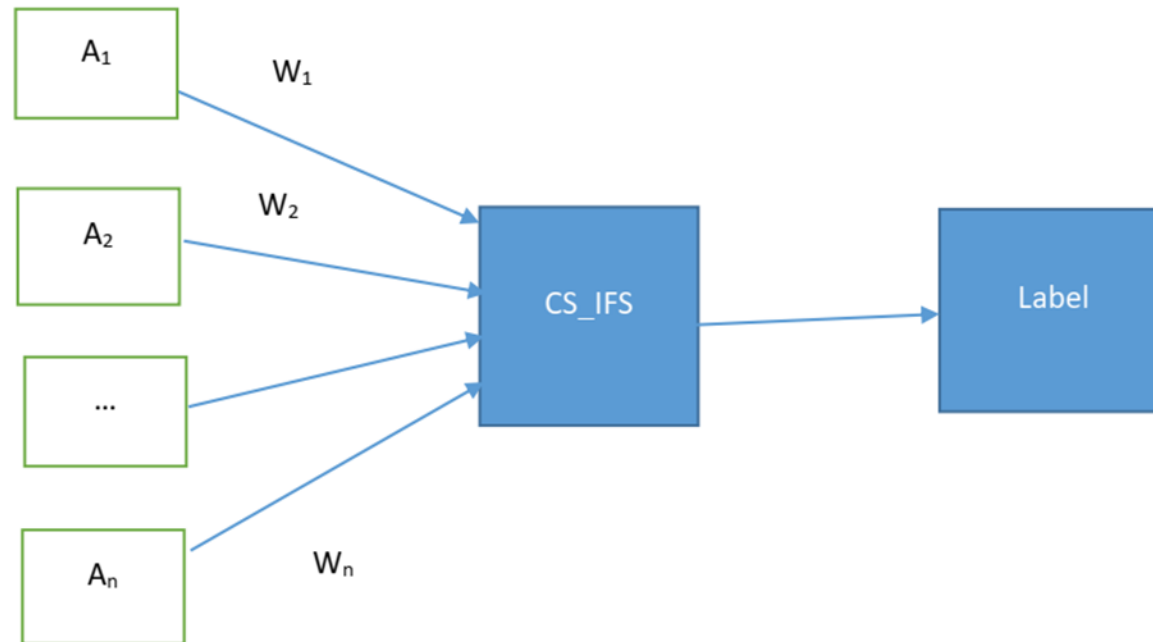
Tổng quát

Bài toán tổng quát: Có m đối tượng $\{O_1, O_2, \dots, O_m\}$ được gán nhãn từ L_1 đến L_k dựa vào n thuộc tính A_1 đến A_n .

Đối tượng	Thuộc tính (Inputs)					Lớp (nhãn)
	A_1	A_2	A_3	A_n	
O_1	r_{11}	r_{1n}	L_1
O_2	r_{21}		r_{2n}	L_2
O_3	r_{31}	r_{3n}	L_1
O_4	r_{41}		r_{4n}	L_3
O_5	r_{51}		r_{5n}	L_2
O_6	r_{61}		r_{6n}	L_k
...	
O_m	r_{m1}				r_{mn}	L_4

r_{ij} : là các giá trị số thực

Mô hình đề xuất



Thuật toán CS-IFS

Thuật toán CS-IFS

- Bước 1: Mờ hóa dữ liệu: chuyển r_{ij} thành (y_{ij}, n_{ij})

$$y_{ij} = \frac{r_{ij} - \min_{i=1,2,\dots,m} r_{ij}}{\max_{i=1,2,\dots,m} r_{ij} - \min_{i=1,2,\dots,m} r_{ij}} \quad (1)$$

+ Hàm thuộc

+ Hàm không thuộc: $n_{ij} = \frac{1 - y_{ij}}{1 + y_{ij}}$ (2) (Surgeno)

Bước 2: Tính các weight (trên tập dữ liệu train) $w_j^{(t+1)} = \frac{w_j^{(t)} T_j}{\sum_{j=1}^n w_j^{(t)} T_j}$ (3) với

$$T_j = \frac{\sum_{i=1}^m |S(y_{ij}, n_{ij}) - S(n_{ij}, y_{ij})|}{m} \quad (4) \text{ và}$$

$$S((y_{ij}, n_{ij})) = \frac{[3 + 2y_{ij} + (y_{ij})^2 - n_{ij} - 2(n_{ij})^2] e^{2y_{ij} - 2n_{ij} - 2}}{6} \quad (5)$$

ở đây t là vòng lặp thứ t ($t=0,1,2,\dots$), $w_j^{(0)} = \frac{1}{n}$

Thuật toán CS-IFS

Bước 3: Tính các tâm của mỗi lớp (cụm) từ tập dữ liệu:

$$S_{L_p} = \left\{ (C_j, r_j^p) \mid j = 1, 2, \dots, n \right\} \left(r_j^p = \frac{1}{|L_p|} \sum_{q=1}^{|L_p|} r_{ij}, \quad p = 1, 2, \dots, k \right)$$

+ Mờ hóa các tâm: $S_{L_p} = \left\{ (C_j, (y_j^{L_p}, n_j^{L_p})) \mid j = 1, 2, \dots, n \right\}$ by using

$$y_j^{L_p} = \frac{r_j^p - \min_{i=1,2,\dots,m} r_{ij}}{\max_{i=1,2,\dots,m} r_{ij} - \min_{i=1,2,\dots,m} r_{ij}}, \quad n_j^{L_p} = \frac{1 - y_j^{L_p}}{1 + y_j^{L_p}}, \quad \text{for all } p = 1, 2, \dots, k.$$

Bước 4. Tính khoảng cách từ mỗi đối tượng đến tâm của các cụm:

$$d(S_i, S_{L_p}) = \sum_{j=1}^n w_j^{(*)} \left| S_i(y_{ij}, n_{ij}) - S_{L_p}(y_j^{L_p}, n_j^{L_p}) \right| \quad (6)$$

for all $p = 1, 2, \dots, k$, $i = 1, 2, \dots, m$

Thuật toán CS-IFS

Bước 5: For all $i=1,2,\dots,m$, we put S_i belong the class $S_{L_{p^*}}$ such that

$$d(S_i, S_{L_{p^*}}) = \min_{p=1,2,\dots,k} \{d(S_i, S_{L_p})\}$$

Bước 6: Tính độ chính xác: $acc_T^t = \frac{n_T^t}{N}$ trong đó n_T^t là tổng số phần tử được phân lớp đúng và N là tổng số phần tử của tập test.

Kết quả chạy mô hình

↳ number of class: 4

len: 145

Check_Result of classification: [1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1,

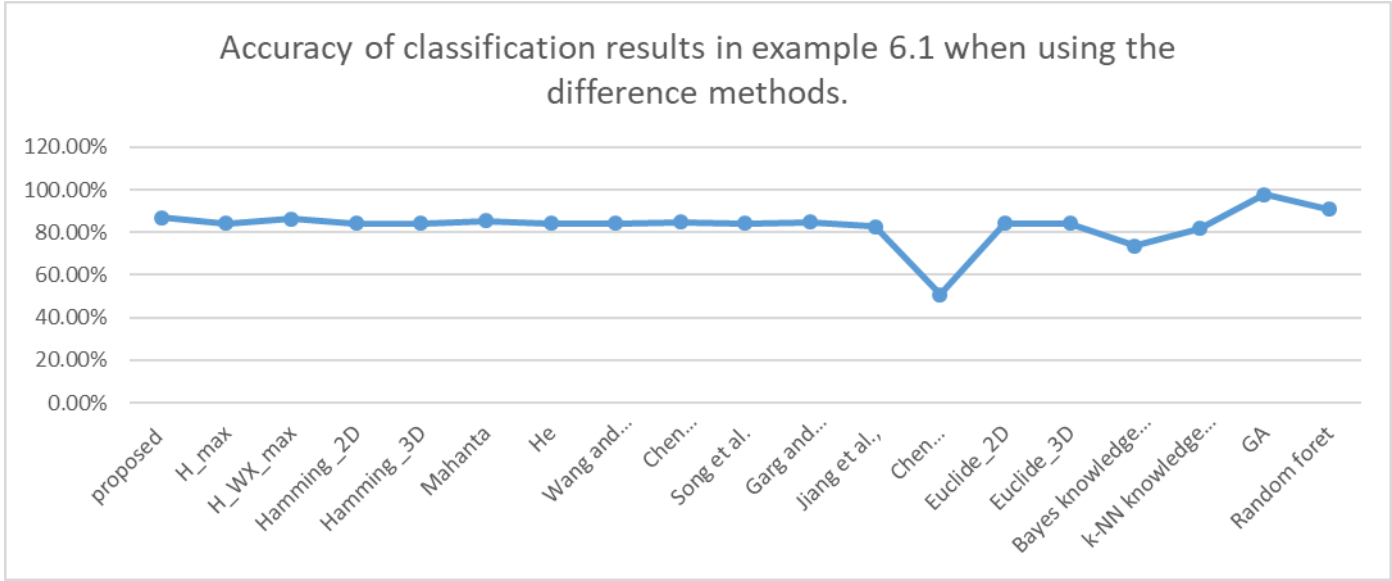
True: 126

Accuracy: 0.869

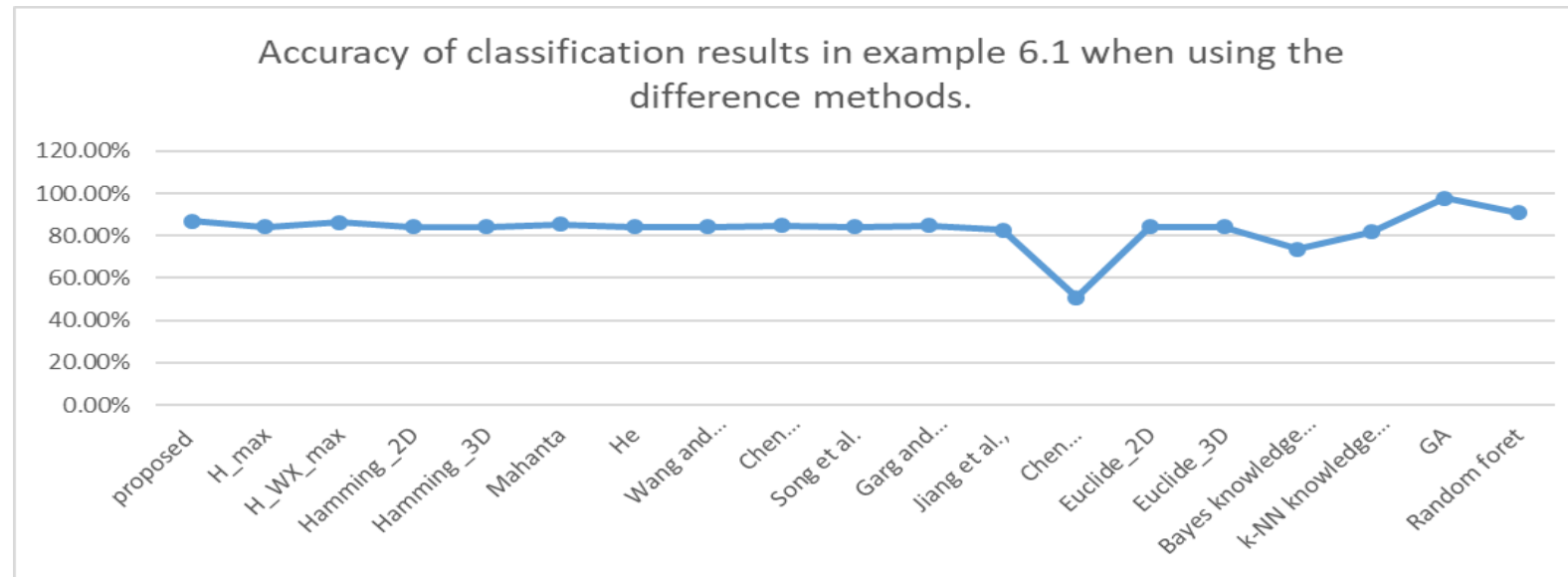
No.	Result				
	Train	Test	Train	Validate	Test
1	81.0078	86.8966	83.4146	83.0189	83.4483
	0.0269	0.0199	0.8560	0.8560	0.0220
2			80.9756	84.9057	82.7586
			3.9925	3.9925	0.0189
3			83.4146	88.6792	86.8966
			0.8508	0.8508	0.0219
4			81.9512	81.1321	78.6207
			1.4324	1.4324	0.0229
5			81.4634	94.3396	81.3793
			0.9466	0.9466	0.0209

No.	Attributes	Result
1	STG,SCG,STR,LPR,PEG	Rate: 94.48275862068965 Time: 0.171067476272583 Feature Importances: STG: 0.0840966177075364 SCG: 0.07695526869228603 STR: 0.05592931054885202 LPR: 0.1677576567499081 PEG: 0.6152611463014175
2	LPR,PEG	Rate: 95.86206896551724 Time: 0.10665225982666016 Feature Importances: LPR: 0.25522854236291764 PEG: 0.7447714576370824
3	PEG	Rate: 84.13793103448276 Time: 0.11164212226867676 Feature Importances: PEG: 1

Kết quả chạy mô hình



So sánh kết quả với một số mô hình khác



Proposed	H_max	H_WX_max	Hamming_2D	Hamming_3D	Mahanta	He	Wang and Xin	Chen and Deng
86.90%	84.14%	86.21%	84.14%	84.14%	85.52%	84.14%	84.14%	84.83%
Garg and Rani	Jiang et al.,	Chen and Deng	Euclide_2D	Euclide_3D	Bayes knowledge classifier	k-NN knowledge classifier	GA	Random forest
84.83%	82.76%	50.86%	84.14%	84.14%	73.80%	81.90%	97.90%	91%

Trên tập dữ liệu OCCUPANCY

date	date	Temperat	Humidity	Light	CO2	HumidityF	Occupancy
1	2/4/2015 17:51	23.18	27.272	426	721.25	0.004793	1
2	2/4/2015 17:51	23.15	27.2675	429.5	714	0.004783	1
3	2/4/2015 17:53	23.15	27.245	426	713.5	0.004779	1
4	2/4/2015 17:54	23.15	27.2	426	708.25	0.004772	1
5	2/4/2015 17:55	23.1	27.2	426	704.5	0.004757	1
6	2/4/2015 17:55	23.1	27.2	419	701	0.004757	1
7	2/4/2015 17:57	23.1	27.2	419	701.6667	0.004757	1
8	2/4/2015 17:57	23.1	27.2	419	699	0.004757	1
9	2/4/2015 17:58	23.1	27.2	419	689.3333	0.004757	1
10	2/4/2015 18:00	23.075	27.175	419	688	0.004745	1
11	2/4/2015 18:01	23.075	27.15	419	690.25	0.004741	1
12	2/4/2015 18:02	23.1	27.1	419	691	0.004739	1
13	2/4/2015 18:03	23.1	27.16667	419	683.5	0.004751	1
14	2/4/2015 18:04	23.05	27.15	419	687.5	0.004734	1
15	2/4/2015 18:04	23	27.125	419	686	0.004715	1
16	2/4/2015 18:06	23	27.125	418.5	680.5	0.004715	1
17	2/4/2015 18:07	23	27.2	0	681.5	0.004728	0
18	2/4/2015 18:08	22.945	27.29	0	685	0.004728	0

Temperat	Humidity	Light	CO2	HumidityF	Occupancy
21.76	31.13333	437.3333	1029.667	0.005021	1
21.79	31	437.3333	1000	0.005009	1
21.7675	31.1225	434	1003.75	0.005022	1
21.7675	31.1225	439	1009.5	0.005022	1
21.79	31.13333	437.3333	1005.667	0.00503	1
21.76	31.26	437.3333	1014.333	0.005042	1
21.79	31.1975	434	1018.5	0.005041	1
21.79	31.39333	437.3333	1018.667	0.005073	1
21.79	31.3175	434	1022	0.00506	1
21.79	31.46333	437.3333	1027.333	0.005084	1
21.79	31.525	437.75	1047.75	0.005094	1
21.79	31.575	441.75	1049	0.005102	1
21.79	31.395	442	1061.5	0.005073	1
21.79	31.3925	441.75	1049	0.005073	1
21.79	31.5	441.5	1048	0.00509	1
21.815	31.5	438	1049.25	0.005098	1
21.815	31.4725	449.5	1051.25	0.005093	1
21.89	31.6	449.5	1060.5	0.005138	1
21.79	31.55	449.5	1059.5	0.005098	1
21.82333	31.73	447.6667	1072	0.005138	1
21.865	31.7225	449.5	1072.333	0.00515	1
21.89	31.745	449.5	1075.333	0.005162	1
21.89	31.7675	449.5	1087	0.005165	1
21.89	31.7675	437.75	1083.25	0.005165	1
21.89	31.79	442.6667	1086.333	0.005169	1
21.89	31.79	437.3333	1091	0.005169	1
21.89	31.89	439	1100	0.005185	1
21.89	31.89	444	1101	0.005185	1
21.89	31.865	444	1107	0.005181	1
21.89	31.79	444	1110	0.005169	1
21.89	31.865	444	1101.75	0.005181	1
21.89	31.73	444	1100	0.005159	1
21.89	31.7225	444	1092.75	0.005158	1
21.89	31.7	444	1092.5	0.005154	1
21.89	31.695	436.5	1094	0.005153	1
21.89	31.56667	434	1085.667	0.005132	1
21.89	31.55	436.5	1047	0.00513	0
21.89	31.36	434	1031	0.005099	0
21.89	31.125	432.75	977.5	0.00506	0

Trên tập dữ liệu OCCUPANCY

No.	Result						
	Train	Test 1	Test 2	Train	Validate	Test 1	Test 2
1	97.6790	97.8236	98.9233	98.6798	99.0792	94.4841	99.2104
	0.3981	0.1985	0.5108	2.9505	2.9505	0.1894	0.5296
2				98.7872	98.9564	89.7936	99.3335
				2.8918	2.8918	0.1459	0.4673
3				96.3617	98.7723	87.9550	99.0156
				2.7012	2.7012	0.1476	0.4522
4				98.0196	98.5881	97.8236	99.0361
				2.7178	2.7178	0.1883	0.4668
5				98.8486	98.4653	95.1220	99.2002
				2.7138	2.7138	0.1473	0.4597

No.	Attributes	Result 1	Result 2
1	Temperature, Humidity, Light, CO2, HumidityRatio	Rate: 94.59662288930582 Time: 0.9436440467834473 Feature Importances: Temperature: 0.07407358049906027 Humidity: 0.02076621548919251 Light: 0.6192129301882042 CO2: 0.2555437410060837 HumidityRatio: 0.030403532817459316	Rate: 97.10828547990155 Time: 1.2820539474487305 Feature Importances: Temperature: 0.11762515573226466 Humidity: 0.024773284895890275 Light: 0.6311773683594154 CO2: 0.1951629714309017 HumidityRatio: 0.031261219581528
2	Light, CO2	Rate: 94.14634146341463 Time: 0.855431318283081 Feature Importances: Light: 0.608854022367372 CO2: 0.3911459776326281	Rate: 96.1033634126333 Time: 1.1552131175994873 Feature Importances: Light: 0.6396158289718371 CO2: 0.3603841710281629

Trên tập dữ liệu IRIS

C1	C2	C3	C4	Class
5.4	3.7	1.5	0.2	Iris-setosa
4.8	3.4	1.6	0.2	Iris-setosa
4.8	3	1.4	0.1	Iris-setosa
4.3	3	1.1	0.1	Iris-setosa
5.8	4	1.2	0.2	Iris-setosa
5.7	4.4	1.5	0.4	Iris-setosa
5.4	3.9	1.3	0.4	Iris-setosa
5	2	3.5	1	Iris-versicolor
5.9	3	4.2	1.5	Iris-versicolor
6	2.2	4	1	Iris-versicolor
6.1	2.9	4.7	1.4	Iris-versicolor
5.6	2.9	3.6	1.3	Iris-versicolor
6.7	3.1	4.4	1.4	Iris-versicolor
6.4	3.2	5.3	2.3	Iris-virginica
6.5	3	5.5	1.8	Iris-virginica
7.7	3.8	6.7	2.2	Iris-virginica
7.7	2.6	6.9	2.3	Iris-virginica

No.	Result				
	Train	Test	Train	Validate	Test
1	90.8333	93.3333	91.6667	95.8333	93.3333
	0.0340	0.0329	0.5939	0.5939	0.0260
2			88.5417	100.0000	93.3333
			0.6790	0.6790	0.0289
3			90.6250	91.6667	96.6667
			0.9509	0.9509	0.0299
4			92.7083	100.0000	93.3333
			0.7247	0.7247	0.0279
5			91.6667	95.8333	93.3333
			0.7636	0.7636	0.0289

No.	Attributes	Result
1	C1,C2, C3,C4	Rate: 96.6666666666667 Time: 0.12154603004455566 Feature Importances: C1: 0.10185933102048822 C2: 0.026068724031321485 C3: 0.4005835147362672 C4: 0.47148843021192316
2	C3,C4	Rate: 96.6666666666667 Time: 0.1198735237121582 Feature Importances: C3: 0.5069662629054572 C4: 0.4930337370945428

Tập dữ liệu spam (57 thuộc tính)

No.	Result				
	Train	Test	Train	Validate	Test
1	61.2500	61.2378	79.0082	89.8098	61.9978
	0.1879	0.1064	45.1500	45.1500	0.1445
2	61.2772	61.2378	88.5530	89.4022	82.0847
	0.1781	0.1106	132.8890	132.8890	0.1196
3	61.2500	61.1292	83.0978	88.9946	67.3181
	0.1931	0.0941	65.5977	65.5977	0.1196
4	61.3587	60.9121	88.9946	91.5761	88.9251
	0.1777	0.0996	61.3784	61.3784	0.0982
5	61.1141	61.8893	62.7378	86.5489	86.7535
	0.1872	0.0986	53.3400	53.3400	0.0810

No.	Attributes	Result
1	All	Rate: 96.30836047774159 Time: 0.5831830501556396