The Application of Fuzzy C-Means Clustering to Sea-Ice Data Analysis

Naxin Chen, Ying Ren

Department of Applied Mathematics
Dalian Maritime University
Dalian China P. R.

Abstract –

Sea-ice blocks classification is one of the most important aspects in various sea-ice researches, especially when the purpose of which is to find out how to protect engineering structures against impact of ice. For the engineering purpose, sea-ices floating on a given area are usually classified into a few groups mainly according to two indices, diameter and thickness. The Fuzzy *C*-Means (FCM) is one of the algorithms for clustering, which is suitable to cluster large data with fuzzy boundary between every two. We apply The Fuzzy *C*-Means clustering algorithm in the classification of a data set made of 460 sea-ice measurements. The data set finally is classified into eight groups. Results show that The Fuzzy *C*-Means is suitable to be used in clustering a large group of sea-ice measurements into a few sub-groups, which will be considered in the design of protecting facilities to avoid sea-ice's impact on engineering structures.

Index Terms - Clustering, FCM, sea-ice

I. THE PURPOSE OF SEA-ICE CLASSICATION

There are rich petroleum and natural gas resources in sea area around the world. The exploration and exploitation of oil and natural gas from wells drilled in sea area are much more complicated than in the land. Oceanic disaster is one of the main factors that result in the complexity. For instance, while in South China Sea the most serious oceanic disaster is typhoon, in Bohai area, sea-ice is the most destroyed one, especially in the northern area, which is called Liaodong Bay. This sea area is covered by dense sea-ice pieces, thin or thick and large or small, during a long period, which is from the middle of December of the year to the end of February of the next year. These floating sea-ices, particularly the ones with both quality and strength on a certain degree, seriously threaten the safety of the engineering structures built in the sea area. Therefore, necessary measures should be taken in order to protect the structures against the impact from the sea-ices.

There are several measures designed to avoid the oil exploiting facilities from sea-ice disaster. Most of them are based on increasing the strength of structure itself, which undoubtedly leads to a drastic increase of cost. Therefore, other measures have been being tried recently. One alternative solution is to build another structure, facing the main moving direction of the sea-ices, besides the main structure. The major function of the affiliated structure is to destroy lots of sea-ices with a destructive momentum. In other word, the affiliated structure damages those dangerous sea-ices into some small safe ones. Obviously, the very first step of designing the affiliated structure should obtain the statistics of the sea-ices sizes in the sea area where the engineering structure locates. The second step should classify all the sea-ices into a few groups according to their diameter and thickness, both of which are essential in determining the quality and strength of sea-ices. Finally, the third one should make a blueprint first, and further discussed details of the design according to the data obtained above. In summary, sea-ices classification takes an important part in the design of against-ice measures.

Considering the characteristics of sea-ices itself and the chief aim of classification, the classification should be done according to the two indices of sea-ice, diameter (or area) and thickness, instead of to only one of them or to a new index based on them, such as the volume of sea-ices. Actually, large but very thin sea-ices can't damage structures seriously because of the strength of the ices, neither can thick but very small ones because of the quality. Meanwhile, the obtained statistics show that the boundary between the features of sea-ice can't be easily distinguished. So, certain sea-ices do not completely belong to a single class, but partially belong to the other classes too. For the reasons discussed above, the fuzzy clustering method provides a better method to classify sea-ices, though there are various clustering methods available.

II. FUZZY C-MEANS CLUSTERING

Given a set of data, Fuzzy C-Means clustering (FCMC) performs clustering by iteratively searching for a set of fuzzy partitions and the associated clustering centers that represent the structure of the data as best as possible. The FCMC algorithm relies on the user to specify the number of clusters present in the set of data to be clustered. Given the number of clusters c, FCMC partitions the data $X=\{x_1, x_2, \cdots, x_n\}$ into c fuzzy partitions by minimizing the within group sum of squared error objective function as follows (eqn 2.1).

$$J_m(U,V)=\sum_{k=1}^{n}\sum_{i=1}^{c}(U_{ik})^m\|x_k-v_i\|^2 \quad 1\le m\le\infty \qquad \text{eqn 2.1}$$

where $J_m(U,V)$ is the sum of squared error for the set of fuzzy clusters represented by the membership matrix U, and the associated set of cluster centers V. $\|\bullet\|$ is some inner product-induced norm. In the formula, $\|X_k-V_i\|^2$ represents the distance between the data X_k and the cluster center V_i. The squared error is used as a performance index that measures the weighted sum of distances between cluster centers and elements in the corresponding fuzzy clusters. The number m governs the influence of membership grades in the performance index. The partition becomes fuzzier with increasing m and it is proven that the FCMC algorithm converges for any $m\in(1,\infty)$ [I]. The necessary conditions for eqn 2.1 to reach its minimum are Eqn 2.2

$$U_{ik}=\left(\sum_{j=1}^{c}\left(\frac{\|x_k-v_i\|}{\|x_k-v_j\|}\right)^{2/(m-1)}\right)^{-1} \quad \forall i,\forall k \qquad \text{Eqn 2.2}$$

And Eqn 2.3

$$v_i=\frac{\sum_{k=1}^{n}(U_{ik})^m x_k}{\sum_{k=1}^{n}(U_{ik})^m} \qquad \text{Eqn 2.3}$$

In each iteration of the FCMC algorithm, matrix U is computed using eqn 2.2 and the associated cluster centers are computed as eqn 2.3. This is followed by computing the square error in eqn 2.1. The algorithm stops when either the error is below a certain tolerance value or its improvement over the previous iteration is below a certain threshold.

III. THE CLASSIFICATION OF SEA-ICE DATA

The given data set consists of 460 measurements obtained a sea area in Bohai. The data, which are given in Table 1, represented as vectors in a 2-D measurement space, in which two variables are diameter and thickness of sea-ice. With the purpose of providing data for the design of anti-ice facilities, the number of clusters should be in the range of possibilities, which is from 4 to 10. Within the range of possibilities, we classify the data set into the corresponding groups, respectively. Furthermore, we assess the clustering results with FOM and PC, and the cluster validity indices are given in Table 2.

Table 2 Validity indices for different for the data

cluster	FOM	PC
4	0.1441	0.6315
5	0.1241	0.5929
6	0.1234	0.5157
7	0.1095	0.5373
8	0.1041	0.5317
9	0.0997	0.5208
10	0.0926	0.5176

Considering both the application of the clustering results and their validity, the data set is suggested to be classified into 8 groups. The result is illustrated in Table 3.

Table 3 clustering result using c=8

cluster	center		Number of data included in each cluster
	D	TH	
I	91.18	24.58	70 (5 8 17 22 36 39 63 68 71 83 93 95 105 114 116 123 124 127 131 132 136 139 141 145 157 164 167 171 175 178 179 187 191 192 199 206 217 228 229 247 249 257 265 266 282 287 290 291 295 313 314 322 324 326 330 331 333 345 348 358 360 364 367 372 373 396 419 428 431 446)
II	216.33	24.87	42(2 10 13 53 56 72 73 75 76 81 102 106 113 118 119 142 154 165 180 188 201 215 219 222 262 264 279 294 300 304 306 318 328 329 334 336 378 383 410 411 418 427)
III	213.31	12.73	51 (3 4 21 25 49 64 70 77 86 108 110 121 126 150 153 170 173 182 194 197 230 238 239 240 243 251 252 258 268 286 298 309 319 321 335 337 340 355 356 359 374 376 379 386 398 417 436 445 447 448 451)
IV	353.87	16.04	36 (1 14 15 20 23 34 57 79 99 115 133 143 144 151 152 176 220 271 278 305 310 312 317 332 353 405 423 433 437 441 449 452 453 455 456 457)
V	139.46	17.98	75(19 30 40 55 60 61 65 67 80 87 92 97 98 100 101 111 112 125 129 155 156 166 169 174 181 184 186 190 200 203 207 211 213 216 218 227 232 233 235 236 241 242 248 250 256 269 277 280 293 301 302 303 316 327 341 342 344 346 349 350 352 354 361 366 368 375 377 389 393 407 415 438 443 444 454)
VI	250.50	7.27	53(24 27 29 31 37 38 47 48 54 74 82 85 88 104 128 135 140 147 158 159 161 168 177 189 198 204 208 214 224 226 231 234 237 254 263 267 270 288 297 343 365 380 397 399 403 408 414 420 422 439 442 458 460)
VII	101.88	14.30	51(12 28 43 66 84 90 91 94 103 107 120 122 160 172 193 202 205 209 210 221 225 253 255 260 272 274 281 283 292 296 308 315 320 325 338 339 347 362 363 371 381 382 390 391 400 402 404 412 426 432 435)
VIII	108.54	7.25	82(6 7 9 11 16 18 26 32 33 35 41 42 44 45 46 50 51 52 58 59 62 69 78 89 96 109 117 130 134 137 138 146 148 149 162 163 183 185 195 196 212 223 244 245 246 259 261 273 275 276 284 285 289 299 307 311 323 351 357 369 370 384 385 387 388 392 394 395 401 406 409 413 416 421 424 425 429 430 434 440 450 459)

Table 1 Data set

No.	D	TH	No.	D	TH	No.	D	TH	No.	D	TH	No.	D	TH	No.	D	TH	No.	D	TH	No.	D	TH	No.	D	TH
1	390.11	15.37	53	227.80	22.61	105	61.11	27.28	157	45.65	24.70	209	85.66	15.37	261	103.97	7.32	313	88.78	23.93	365	231.95	7.60	417	202.32	13.46
2	229.46	26.25	54	294.41	5.01	106	174.94	25.99	158	193.32	7.88	210	97.67	13.46	262	248.00	30.09	314	58.05	22.08	366	162.75	17.00	418	239.74	22.61
3	257.29	10.68	55	149.59	17.00	107	58.14	15.37	159	193.32	7.88	211	184.89	15.64	263	548.62	6.74	315	88.91	13.46	367	128.35	26.51	419	136.17	27.54
4	198.29	15.64	56	186.92	22.87	108	187.21	14.28	160	96.30	13.46	212	107.45	7.32	264	191.72	27.54	316	190.57	15.64	368	188.77	16.19	420	213.11	5.01
5	71.85	25.73	57	347.51	13.46	109	68.38	6.74	161	192.28	3.86	213	198.29	16.46	265	48.79	21.28	317	324.37	18.90	369	86.21	4.43	421	93.06	8.44
6	108.98	8.72	58	130.27	8.72	110	173.05	10.68	162	118.49	7.03	214	222.04	6.16	266	48.79	21.28	318	200.31	28.32	370	93.46	4.43	422	242.89	6.74
7	159.42	5.88	59	173.54	2.99	111	80.57	18.63	163	119.37	7.03	215	219.48	21.28	267	277.91	10.12	319	145.19	11.52	371	148.67	15.64	423	330.15	17.28
8	149.27	23.93	60	156.13	17.00	112	79.71	16.73	164	53.89	27.54	216	191.37	15.64	268	162.05	10.12	320	95.48	13.46	372	119.60	23.93	424	94.13	8.72
9	91.08	10.40	61	69.53	18.09	113	225.00	23.14	165	310.67	30.09	217	58.05	22.08	269	61.24	18.63	321	234.40	14.28	373	120.35	23.40	425	93.06	8.44
10	394.43	27.03	62	93.06	8.44	114	67.17	25.73	166	122.37	18.90	218	174.09	16.19	270	193.32	7.88	322	109.34	23.67	374	297.02	11.24	426	74.88	14.55
11	93.06	8.44	63	136.09	24.44	115	323.97	15.37	167	107.78	26.51	219	225.59	24.96	271	718.06	8.16	323	122.01	7.03	375	154.67	17.00	427	160.57	24.96
12	153.84	14.28	64	154.12	10.40	116	88.06	23.67	168	231.15	7.32	220	813.38	19.16	272	133.72	15.10	324	83.47	27.54	376	214.34	14.55	428	128.74	25.22
13	448.18	25.73	65	107.54	19.69	117	50.65	9.00	169	108.12	17.00	221	93.29	13.46	273	136.24	10.40	325	107.81	16.19	377	88.36	20.22	429	76.59	5.88
14	305.49	14.28	66	103.37	12.92	118	169.26	23.93	170	215.13	11.80	222	201.00	27.54	274	148.08	13.74	326	122.51	24.19	378	221.57	25.22	430	80.12	5.88
15	330.18	15.92	67	153.46	17.00	119	433.11	26.25	171	53.69	23.40	223	86.72	8.16	275	124.65	7.03	327	66.34	20.49	379	233.22	12.92	431	130.21	26.25
16	82.40	8.16	68	72.73	23.67	120	138.08	11.80	172	114.59	15.92	224	222.35	9.56	276	131.69	7.03	328	164.65	23.67	380	276.71	8.72	432	95.72	14.55
17	114.26	25.73	69	82.84	5.88	121	204.77	12.92	173	183.11	10.40	225	126.67	12.08	277	163.45	18.09	329	219.85	24.70	381	94.55	14.01	433	361.80	21.28
18	93.06	8.44	70	179.02	9.56	122	60.61	15.37	174	123.66	17.00	226	193.32	7.88	278	387.29	19.16	330	74.07	24.44	382	94.55	14.01	434	79.84	5.88
19	126.31	19.96	71	73.64	23.14	123	112.91	22.61	175	141.58	23.93	227	120.06	20.22	279	269.45	23.40	331	120.53	25.22	383	187.21	23.67	435	85.56	14.55
20	663.76	14.28	72	261.55	24.44	124	110.59	23.67	176	372.72	12.36	228	140.81	22.61	280	144.97	21.02	332	392.28	10.96	384	93.06	8.44	436	179.32	11.80
21	269.85	11.24	73	185.63	22.61	125	208.40	17.82	177	184.49	6.16	229	63.45	25.73	281	82.26	15.37	333	137.58	21.81	385	85.61	8.16	437	378.21	19.96
22	79.14	24.44	74	193.32	7.88	126	208.19	12.92	178	45.30	22.34	230	208.81	12.36	282	113.89	22.08	334	192.12	25.73	386	297.02	11.24	438	157.38	18.37
23	252.72	17.82	75	190.91	22.61	127	38.49	23.67	179	43.46	23.67	231	309.62	7.03	283	91.10	13.46	335	187.60	14.28	387	89.73	8.16	439	400.73	7.32
24	263.49	6.16	76	336.97	21.55	128	260.75	5.59	180	251.33	24.44	232	199.03	18.63	284	109.93	7.32	336	159.43	25.99	388	95.30	8.72	440	165.62	7.60
25	214.34	10.40	77	236.44	15.64	129	162.39	20.75	181	165.86	21.02	233	117.14	16.46	285	106.41	7.32	337	193.92	12.92	389	123.56	17.00	441	302.78	14.01
26	79.59	5.88	78	93.06	8.44	130	132.57	7.03	182	222.29	10.40	234	207.32	8.44	286	213.32	12.92	338	94.55	14.01	390	94.55	14.01	442	324.14	7.03
27	213.80	8.16	79	435.86	11.52	131	51.43	24.19	183	81.32	7.32	235	73.39	18.63	287	73.38	21.02	339	148.08	13.74	391	61.24	15.64	443	104.14	16.73
28	130.47	11.52	80	128.88	20.22	132	105.95	27.03	184	174.15	16.73	236	129.03	18.90	288	217.64	7.88	340	202.41	15.37	392	93.06	8.44	444	156.68	17.00
29	307.15	10.12	81	241.51	26.25	133	280.87	16.46	185	121.13	7.03	237	265.50	7.88	289	104.01	10.40	341	120.51	17.00	393	116.78	20.75	445	180.61	11.80
30	91.31	18.63	82	284.53	3.57	134	128.17	7.03	186	131.25	18.90	238	192.73	15.64	290	48.79	21.28	342	142.08	17.00	394	92.11	4.43	446	132.54	31.35
31	435.14	6.45	83	83.34	24.96	135	207.63	6.74	187	91.60	27.54	239	160.21	12.64	291	88.52	27.54	343	197.12	4.15	395	84.09	5.88	447	254.23	10.96
32	93.06	8.44	84	119.62	11.80	136	87.43	21.55	188	161.00	23.67	240	249.78	15.37	292	148.08	13.74	344	148.03	17.00	396	71.85	24.19	448	183.57	11.80
33	81.23	5.88	85	191.44	5.01	137	127.29	7.03	189	240.63	5.88	241	169.92	20.49	293	93.40	18.37	345	73.64	23.14	397	258.47	7.88	449	392.11	12.36
34	352.03	15.64	86	236.44	15.64	138	130.81	7.03	190	239.74	20.22	242	124.59	18.90	294	218.21	25.22	346	119.31	18.63	398	233.22	12.92	450	113.77	6.16
35	98.96	8.72	87	158.83	17.00	139	61.11	27.28	191	118.82	22.34	243	231.01	12.92	295	83.34	23.40	347	94.55	14.01	399	221.72	8.72	451	181.90	11.80
36	78.67	23.67	88	350.30	7.60	140	193.32	7.88	192	101.83	22.08	244	126.41	7.03	296	88.48	14.55	348	110.45	27.28	400	124.48	12.08	452	307.68	14.55
37	232.20	8.16	89	128.24	7.32	141	58.91	25.48	193	79.06	15.37	245	133.45	7.03	297	207.63	6.74	349	160.73	17.00	401	93.47	8.72	453	349.76	20.49
38	199.72	9.56	90	94.55	14.01	142	162.63	21.81	194	248.25	11.24	246	104.01	10.40	298	203.06	12.92	350	104.14	16.73	402	122.42	14.55	454	255.38	18.63
39	125.13	22.34	91	94.55	14.01	143	294.37	17.82	195	104.01	7.32	247	84.63	23.14	299	122.89	7.03	351	85.06	8.16	403	333.14	5.30	455	553.43	17.28
40	207.32	17.00	92	120.51	17.00	144	367.57	12.08	196	111.49	7.32	248	194.58	15.92	300	269.45	30.59	352	155.96	17.00	404	94.55	14.01	456	257.15	16.19
41	92.11	4.43	93	45.54	26.77	145	101.04	25.22	197	309.07	12.08	249	63.45	21.02	301	183.19	21.02	353	503.28	23.93	405	619.64	23.67	457	336.44	18.63
42	90.01	4.43	94	94.55	14.01	146	129.93	7.03	198	372.37	5.59	250	174.09	17.00	302	177.64	20.75	354	182.41	21.02	406	78.29	8.16	458	246.87	7.88
43	125.49	15.64	95	45.54	26.77	147	193.32	7.88	199	56.01	26.77	251	163.15	10.12	303	77.14	18.63	355	179.16	13.74	407	121.71	17.00	459	106.98	6.16
44	84.00	3.28	96	100.19	6.16	148	148.13	8.16	200	187.05	15.64	252	163.15	10.12	304	162.63	25.73	356	233.22	12.92	408	190.29	6.74	460	356.35	8.72
45	79.39	8.16	97	80.76	17.55	149	125.53	7.03	201	309.43	22.34	253	164.90	15.10	305	251.33	17.00	357	150.89	9.84	409	143.45	10.12			
46	120.25	7.03	98	152.08	17.00	150	209.90	12.92	202	65.14	15.37	254	308.59	3.57	306	253.60	21.02	358	105.90	28.06	410	151.33	27.54			
47	213.11	5.30	99	653.92	23.40	151	379.53	15.10	203	130.15	18.90	255	88.48	14.55	307	68.38	6.74	359	209.79	12.64	411	218.32	24.96			
48	240.79	6.74	100	76.87	18.37	152	251.33	17.00	204	305.00	8.16	256	189.84	16.19	308	78.67	13.19	360	73.64	23.14	412	133.99	11.80			
49	180.59	10.40	101	174.39	15.92	153	252.97	10.40	205	78.96	15.37	257	128.57	22.87	309	206.48	12.92	361	121.29	18.63	413	95.41	4.43			
50	83.41	8.16	102	221.78	21.55	154	349.72	29.08	206	64.44	25.73	258	211.61	12.92	310	251.33	17.00	362	94.55	14.01	414	280.86	7.32			
51	95.64	8.72	103	86.72	13.46	155	80.10	18.63	207	189.21	15.64	259	123.77	7.03	311	129.05	7.03	363	94.55	14.01	415	97.04	16.73			
52	81.34	5.88	104	463.79	8.16	156	126.81	18.90	208	312.98	7.03	260	82.36	15.37	312	496.36	12.36	364	45.54	26.77	416	100.81	4.43			

In fact, among the eight clusters, ices belonging to cluster II,IV,V should be chiefly considered in the design of ice-broken facilities. These types of ices have enough momentum because of their quality, and enough strength because of their thickness, to impact engineering structures. The ices belonging to cluster I,VII don't possess enough quality due to their small diameter, although with enough strength due to the thickness. And the ices belonging to the rest of the clusters are so thin that they are easily broken into small ones with the impact from circumstance, therefore these types of ices can be ignored only from aspect of the safety of engineering structures.

IV. CONCLUSIONS

In section III, we have identified these ices with similar qualities and strengths (or similar diameters and thickness). It suggests that the Fuzzy *C*-Means clustering algorithm can be used for ices clustering as well as other data analysis in the similar situations. On the other hand, a useful sea-ice distribution pattern is also provided for the design of against-ice facilities.

V. DISCUSSION

This research based on the remote sensing data in a usual ice-year. The research should be made out completely, including the contents described as follows:

1. Research on different sea ice cavern in typical years.

2. Research on the different origination, the floating sea ice originating from the different place has different properties, for example, the floes coming off from the beach or sub-aqueous delta are more thick and large. It is more dangerous for engineering structures.

3. Research on classification in the whole process of the interaction between floes and structures in order to state the inter- impact between floes and the structures.

References

[1] Bezdek J.C., Pattern Recognition with Fuzzy Objective Function Algorithms, 1981, New York: Plenum Press.

[2] Chong, A.,Gedeon,T.D., Koczy, L.T., A Hybrid Approach for Solving the Cluster Validity Problem, 14th International Conference on Digital Signal Processing, Volume 2, 1-3 July 2002.

[3] Benson S. Y. Lam, Hong Yan, "Cluster Validity for DNA Microarray Data Using a Geometrical Index," Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 18-21 August 2005.

[4] The Marine Environment in Offshore Working Area of Liaohe Oil Field, CNPC, the Press of DLMU,1999, Dalian, China P.R.

Biographies: Naxin Chen, born in 1969, Lecturer. Her research interests focus on fuzzy algebras and marine environment. E-mail: cnxdlmu@sina.com.cn