



Strategy for primary processing of social networks data using hierarchy analysis method

O V Chengar, V I Shevchenko, E N Maschenko, D V Moiseev and A S Soina

Strategy for primary processing of social networks data using hierarchy analysis method

Criterion	C1	C2	C3	C4	b	x
C1 Dependence on source data volume	1	3	5	7	3,2011	0,5416
C2 Clusters optimal number	0,333	1	5	6	1,7778	0,3008
C3 Algorithm computational complexity	0,2	0,2	1	5	0,6687	0,1131
C4 Algorithm performance	0,1429	0,1667	0,2	1	0,2627	0,0444
TOTAL B					5,9103	0,9999 ← Normalization Check
y	1,6759	4,3667	11,2	19	λmax 4,3315	CI 0,0829
						CR 9,2111 %

Alternatives	A1	A2	A3	A4	A5	A6	b	z
A1 Weighted pair group average	1	1	0,33	0,2	0,2	0,5	0,4331	0,0567
A2 Unweighted pair group centroid	1	1	0,33	0,2	0,2	0,5	0,4331	0,0567
A3 Ward method	3	3	1	0,25	0,25	0,5	0,8094	0,106
A4 K-means	5	5	4	1	1	2	2,4183	0,3168
A5 G-means	5	5	4	1	1	2	2,4183	0,3168
A6 Minimum spanning tree	2	2	2	0,5	0,5	1	1,1225	0,147
TOTAL B							7,6347	1 ← Normalization Check
y	17	17	11,66	3,15	3,15	6,5	λmax 6,1151	CI 0,023
								CR 1,8548 %

We propose a strategy for investigating clustering analysis methods applied to analyze social network users' communities with the help of the Analytic Hierarchy Process (AHP). One considers the following methods of cluster analysis: Weighted pair group average method, Unweighted pair group centroid method, Ward method, K-means, G-means, Minimum spanning tree. It is planned to examine six alternative methods of cluster analysis (A1-A6).

Strategy for primary processing of social networks data using hierarchy analysis method

	Alternatives	C1	C2	C3	C4
A1	Weighted pair group average	0,0567	0,1129	0,2939	0,1321
A2	Unweighted pair group centroid	0,0567	0,1129	0,1682	0,1304
A3	Ward method	0,106	0,1129	0,1156	0,176
A4	K-means	0,3168	0,3177	0,0642	0,1981
A5	G-means	0,3168	0,3177	0,0642	0,2313
A6	Minimum spanning tree	0,147	0,0259	0,2939	0,1321

x	ISi
0,5416	0,023
0,3008	0,0024
0,1131	0,0142
0,0444	0,0116

Rating criteria

	Alternatives	C1	C2	C3	C4
A1	Weighted pair group average	4	2	1	4
A2	Unweighted pair group centroid	4	2	2	5
A3	Ward method	3	2	3	3
A4	K-means	1	1	4	2
A5	G-means	1	1	4	1
A6	Minimum spanning tree	2	3	1	4

Hierarchy CR 1,7%

	Alternatives	1	2	3	4	5	6
A1	Weighted pair group average	1	1		2		
A2	Unweighted pair group centroid		2		1	1	
A3	Ward method		1	3			
A4	K-means	2	1		1		
A5	G-means	3			1		
A6	Minimum spanning tree	1	1	1	1		

Here you see the results. Based on the calculations made, it can be concluded that the G-means method is the most preferable for use, and the worst alternative for these quality criteria and the presented weight sets is the Weighted pair group average method.

- However, it should be taken into account that a significant factor affecting the quality of the results of this strategy is the initial choice of weights in pairwise comparisons of criteria and alternatives, as a result of which the expert may make an error.
- In this regard, the introduction of an additional method of voting by experts, such as the “Delphi” method, will reduce the risk of incorrect prioritization.

*The work was carried out with the financial support of
Sevastopol State University under the internal grant №.
31/06-31.*

Thank you for attention!