Годишник на секция "Информатика"
Съюз на учените в България
Том 8, 2015-2016, 71-78

Annual of "Informatics" Section
Union of Scientists in Bulgaria
Volume 8, 2015-2016, 71-78

INTERCRITERIA ANALYSIS WITH MISSING DATA: AN EXAMPLE WITH ATMOSPHERIC DATA

Todor Kostadinov

Prof. Dr. Asen Zlatarov University
1 Prof. Y.Yakimov, Organic Department 108, 8010 Burgas, Bulgaria
e-mail: kostadinov_todor@btu.bg

Abstract: In this article the impact of incorrect reading or data lost over the novel approach of Intercriteria Analysis is investigated. The study aims to discover the amount of admissible errors or loss of parameters in single criterion or object and in case of multiple errors in different objects/criteria. The analysis of overall influence over the level of uncertainty and correlation level is carried out.
Keywords: Intercriteria analysis, error estimation, uncertainty level, correlation level.

1 Introduction

In present days the requirements of analysis and assessment [1] over certain fields that involve a decision making process are increasing. For example an analysis of the climate parameters and the harvest of different plant cultures during the past years can provide vital information about the conditions that provide the best results. Some other examples are the assessment of risks and the processes in the economics [2], establishment of different pollution factors and disasters [3], and the factors that caused them and etc. It is of major significance to work with accurate readings due to the fact that incorrect or incomplete readings can result a deviation from the correct decisions, leading to bad results and potential losses. There are two major causes of errors – measurement errors and human errors. While measurement errors can be caused by interference or failure in the equipment modules and are usually constant during the whole reading cycle, the human errors are caused by distraction, neglect or health issues that can cause single error or even partial loss of data, rearrangement and etc.

This paper addresses the issues caused by several types of errors – single errors in one object or criterion, multiple errors in one object /criterion and multiple errors across the whole data set along with the impact over the results obtained by the approach of InterCriteria Analysis – ICA. The application of the ICA over several data sets of the same size, containing equivalent weather and atmosphere related parameters with progressively increasing number of errors aims to discover how different type and amount of errors affect the correlations obtained in the initial error free data set.

2 Intercriteria assessment over atmosphere data

The Intercriteria Analysis is an approach that facilitates the decision making process by providing the correlation of certain criteria for a set of objects. The larger the set of objects and criteria – the better assessment it provides. The arrays of data obtained by the measurement of many objects against many criteria are processed until correlations are calculated for each pair of criteria in the form of intuitionistic fuzzy pairs of values in the [0;1]-interval.

The data [4, 5] is sampled in every three hours from the area of Burgas in the first week of the new year. The set of criteria includes 20 atmospheric parameters as criteria according to the ICA terminology, shown in Table 1.

1 - minimal air temperature	11 - visibility
2 - minimal moisture	12 - atmospheric pressure
3 - maximum air temperature	13 - Azote dioxide
4 - air temperature of dewing	14 - Benzene
5 - temperature during maximum moisture	15 - Carbone oxide
6 - moisture	16 - Ozone
7 - wind direction in degrees	17 - Sulfuric dioxide
8 - wind speed m/s	18 - Hydrogen sulfite
9 - altitude in meters	19 - Styrene
10 - clouds average height	20 - Nano dust particles

Table 1. Atmospheric parameters

The possible expectations for intercriteria correlations are between criteria 6–7, 6–8, 6–11, 6–12, 7–12, 7–17, 8–14, 12–14.

The array of data containing the correlations among different criteria sets resulted by the ICA is represented as a square matrix of intuitionistic fuzzy pairs with dimensions 20x20 rows and pillars, providing the relation between any given criteria in the set. The results are presented in two tables according to the terms of membership – Table 2 and non-membership – Table 3 parts of the fuzzy pairs. The color mapping of the difference between the maximum value of membership and the maximum value of non-membership gradates from green to red. In those terms – a strong consonance is present when the fuzzy pair specifies maximum degree of membership and minimum degree of non membership and vice versa for the strong dissonance.

The geometrical representation of the results in Table 2, 3 is illustrated in Figure 1 where every criteria pair provides the intercriteria correlation.

μ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1,00	0,86	0,79	0,77	0,87	0,63	0,55	0,52	0,60	0,47	0,35	0,29	0,61	0,43	0,44	0,40	0,51	0,35	0,44	0,62
2	0,86	1,00	0,74	0,72	0,82	0,65	0,55	0,42	0,57	0,49	0,32	0,27	0,64	0,43	0,51	0,35	0,46	0,40	0,45	0,55
3	0,79	0,74	1,00	0,70	0,73	0,59	0,51	0,50	0,55	0,46	0,34	0,28	0,60	0,46	0,43	0,38	0,49	0,33	0,40	0,62
4	0,77	0,72	0,70	1,00	0,77	0,69	0,51	0,52	0,58	0,52	0,26	0,22	0,57	0,41	0,48	0,35	0,47	0,37	0,43	0,66
5	0,87	0,82	0,73	0,77	1,00	0,73	0,55	0,48	0,61	0,47	0,24	0,23	0,60	0,48	0,46	0,36	0,50	0,38	0,48	0,62
6	0,63	0,65	0,59	0,69	0,73	1,00	0,50	0,38	0,53	0,46	0,02	0,20	0,56	0,50	0,55	0,27	0,40	0,45	0,50	0,56
7	0,55	0,55	0,51	0,51	0,55	0,50	1,00	0,42	0,47	0,38	0,46	0,38	0,58	0,42	0,46	0,55	0,39	0,41	0,41	0,58
8	0,52	0,42	0,50	0,52	0,48	0,38	0,42	1,00	0,64	0,46	0,58	0,58	0,42	0,48	0,29	0,57	0,54	0,37	0,50	0,54
9	0,60	0,57	0,55	0,58	0,61	0,53	0,47	0,64	1,00	0,49	0,42	0,43	0,48	0,50	0,39	0,47	0,52	0,37	0,53	0,54
10	0,47	0,49	0,46	0,52	0,47	0,46	0,38	0,46	0,49	1,00	0,42	0,51	0,44	0,49	0,48	0,36	0,44	0,48	0,52	0,34
11	0,35	0,32	0,34	0,26	0,24	0,02	0,46	0,58	0,42	0,42	1,00	0,77	0,40	0,47	0,41	0,69	0,46	0,48	0,45	0,40
12	0,29	0,27	0,28	0,22	0,23	0,20	0,38	0,58	0,43	0,51	0,77	1,00	0,34	0,55	0,46	0,61	0,48	0,58	0,53	0,31
13	0,61	0,64	0,60	0,57	0,60	0,56	0,58	0,42	0,48	0,44	0,40	0,34	1,00	0,47	0,53	0,36	0,48	0,42	0,45	0,60
14	0,43	0,43	0,46	0,41	0,48	0,50	0,42	0,48	0,50	0,49	0,47	0,55	0,47	1,00	0,57	0,37	0,51	0,57	0,77	0,43
15	0,44	0,51	0,43	0,48	0,46	0,55	0,46	0,29	0,39	0,48	0,41	0,46	0,53	0,57	1,00	0,31	0,35	0,68	0,54	0,38
16	0,40	0,35	0,38	0,35	0,36	0,27	0,55	0,57	0,47	0,36	0,69	0,61	0,36	0,37	0,31	1,00	0,44	0,36	0,34	0,51
17	0,51	0,46	0,49	0,47	0,50	0,40	0,39	0,54	0,52	0,44	0,46	0,48	0,48	0,51	0,35	0,44	1,00	0,35	0,51	0,53
18	0,35	0,40	0,33	0,37	0,38	0,45	0,41	0,37	0,37	0,48	0,48	0,58	0,42	0,57	0,68	0,36	0,35	1,00	0,59	0,27
19	0,44	0,45	0,40	0,43	0,48	0,50	0,41	0,50	0,53	0,52	0,45	0,53	0,45	0,77	0,54	0,34	0,51	0,59	1,00	0,39
20	0,62	0,55	0,62	0,66	0,62	0,56	0,58	0,54	0,54	0,34	0,40	0,31	0,60	0,43	0,38	0,51	0,53	0,27	0,39	1,00

Table 2: Values for μ_{c_i,c_j}

Fig. 1 shows that the consonances are observed between criteria 1–2, 1–3, 1–4, 1–5, 2–1, 2–3, 2–4, 2–5, 15–18, 14–19, which confirms the expectations. Another strong consonance can be noticed between criteria 6–11, 11–12, 11–13. No uncertainty is observed in the intuitionistic triangle as can be seen in Figure 1. This is due to the fact that the criteria correlate well with each other. Also all readings are error free when applied to the ICA.

v	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0,00	0,13	0,15	0,18	0,12	0,31	0,43	0,44	0,36	0,42	0,63	0,70	0,36	0,55	0,55	0,59	0,34	0,58	0,52	0,34
2	0,13	0,00	0,20	0,23	0,16	0,29	0,43	0,54	0,40	0,40	0,67	0,72	0,34	0,56	0,48	0,64	0,40	0,53	0,52	0,41
3	0,15	0,20	0,00	0,25	0,21	0,32	0,42	0,42	0,37	0,41	0,60	0,67	0,33	0,48	0,51	0,56	0,34	0,58	0,53	0,31
4	0,18	0,23	0,25	0,00	0,18	0,23	0,44	0,41	0,36	0,38	0,69	0,74	0,37	0,55	0,48	0,60	0,37	0,55	0,51	0,28
5	0,12	0,16	0,21	0,18	0,00	0,20	0,43	0,48	0,36	0,42	0,74	0,76	0,37	0,51	0,52	0,63	0,36	0,55	0,48	0,34
6	0,31	0,29	0,32	0,23	0,20	0,00	0,43	0,54	0,39	0,41	0,93	0,74	0,36	0,44	0,39	0,67	0,42	0,45	0,41	0,36
7	0,43	0,43	0,42	0,44	0,43	0,43	0,00	0,54	0,49	0,51	0,52	0,60	0,39	0,56	0,52	0,43	0,46	0,51	0,54	0,38
8	0,44	0,54	0,42	0,41	0,48	0,54	0,54	0,00	0,31	0,43	0,39	0,39	0,54	0,49	0,68	0,40	0,31	0,54	0,44	0,40
9	0,36	0,40	0,37	0,36	0,36	0,39	0,49	0,31	0,00	0,39	0,55	0,54	0,48	0,47	0,58	0,50	0,33	0,55	0,41	0,40
10	0,42	0,40	0,41	0,38	0,42	0,41	0,51	0,43	0,39	0,00	0,47	0,39	0,44	0,40	0,42	0,53	0,34	0,38	0,35	0,55
11	0,63	0,67	0,60	0,69	0,74	0,93	0,52	0,39	0,55	0,47	0,00	0,22	0,58	0,52	0,57	0,29	0,41	0,45	0,51	0,57
12	0,70	0,72	0,67	0,74	0,76	0,74	0,60	0,39	0,54	0,39	0,22	0,00	0,64	0,45	0,53	0,38	0,38	0,35	0,43	0,66
13	0,36	0,34	0,33	0,37	0,37	0,36	0,39	0,54	0,48	0,44	0,58	0,64	0,00	0,51	0,45	0,62	0,37	0,50	0,50	0,35
14	0,55	0,56	0,48	0,55	0,51	0,44	0,56	0,49	0,47	0,40	0,52	0,45	0,51	0,00	0,42	0,62	0,35	0,36	0,20	0,53
15	0,55	0,48	0,51	0,48	0,52	0,39	0,52	0,68	0,58	0,42	0,57	0,53	0,45	0,42	0,00	0,68	0,51	0,25	0,42	0,59
16	0,59	0,64	0,56	0,60	0,63	0,67	0,43	0,40	0,50	0,53	0,29	0,38	0,62	0,62	0,68	0,00	0,43	0,57	0,63	0,46
17	0,34	0,40	0,34	0,37	0,36	0,42	0,46	0,31	0,33	0,34	0,41	0,38	0,37	0,35	0,51	0,43	0,00	0,47	0,33	0,32
18	0,58	0,53	0,58	0,55	0,55	0,45	0,51	0,54	0,55	0,38	0,45	0,35	0,50	0,36	0,25	0,57	0,47	0,00	0,34	0,64
19	0,52	0,52	0,53	0,51	0,48	0,41	0,54	0,44	0,41	0,35	0,51	0,43	0,50	0,20	0,42	0,63	0,33	0,34	0,00	0,55
20	0,34	0,41	0,31	0,28	0,34	0,36	0,38	0,40	0,40	0,55	0,57	0,66	0,35	0,53	0,59	0,46	0,32	0,64	0,55	0,00

Table 3: Values for v_{c_i,c_j}

3 Error estimation

In order to establish the role of errors over the decision making process it is necessary to overview several possible cases:

- A single error or a missing value in one object or criterion per data set if the size is below 100 rows and pillars.
- Multiple errors or missing values, consecutive and not consecutive in objects or criteria in the data set – approximately 10 or more percent of the data.

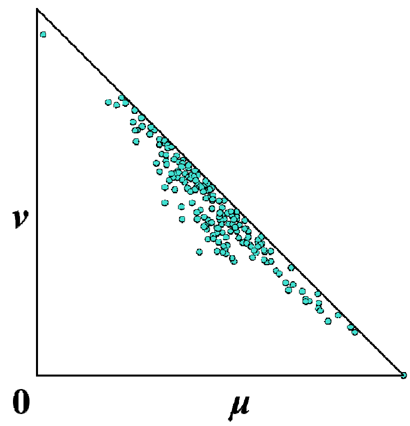
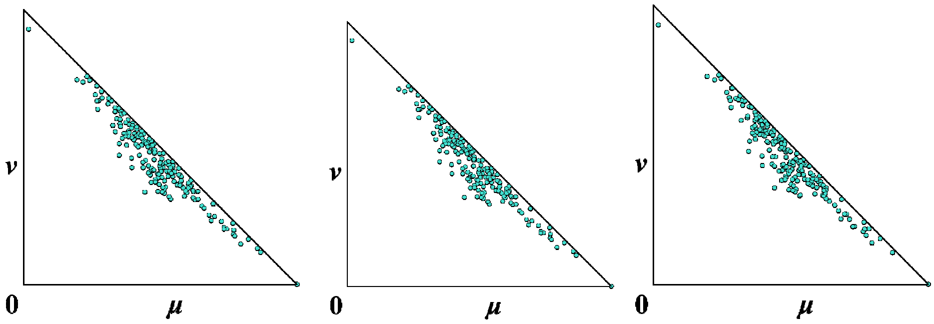


Figure 1. ICA results over the atmospheric data

It is expected that an error or multiple errors will impact the degree of consonance and / or dissonance among the criteria and possible increase of the uncertainty level. The missing values on the other hand would strongly affect the uncertainty level and cause false correlations. To establish an intercriteria analysis scenario over a data set prone to errors it is necessary to compare the results obtained by ICA over data set with presence of errors and over an error free data set. To correctly state the error influence, both data sets must contain exactly the same criteria and objects, but only with and without errors or loss.

3.1 A single error in ICA data

The nature of errors is usually random. That is the reason they can appear in every object and criterion in the data set. The size of the current data set is 20 rows and 57 pillars. It means one error per object / criterion. The error value is simulated as a random number. It can increment or decrement the cell value. To investigate the impact over ICA, the value of criterion number 10 for the object day 3 of January has been reduced by 77 percent.
The results of the altered value in the data set can be observed in Fig. 2, where they are compared with the initial results presented in paragraph 2. No difference can be made in the comparison between the two results. This means that the ICA approach is slightly affected by errors in the data less than 1 per criterion/object of the total amount – Fig. 2 a), b). If the error rate is 1% of all criteria, that means an amount of 13 errors. In Fig. 2 c) no deviation from Figs 2a) and 2b) can be found, which means that ICA is not affected by errors up to 1% of the data.



a) b) c)

Figure 2: Parallel between no errors a) one error b) and 13 errors c) in data set

3.2 Multiple errors in ICA data set

It was proven that accidental errors, peaks and random missing data do not affect the accuracy of the decision making process. Significant errors or data loss can appear due to sensor failure, equipment failure, harsh weather conditions, frosting, abnormal behavior or sabotage. Thus there is a high chance to get consecutive losses in readings or lack of all objects data due to measurement equipment failure. Such event is simulated by removing 10% of the data in several consecutive objects Fig. 3 and several consecutive criteria Fig. 4. It must be specified that there are no more than 12 consecutive measurement missing. The affected criteria are as follow – 1, 2, 4, 7, 9, 12, 16, 18 and 19.

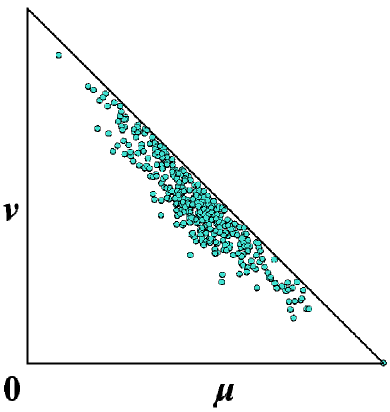


Figure 3: ICA over criteria data with 10% errors

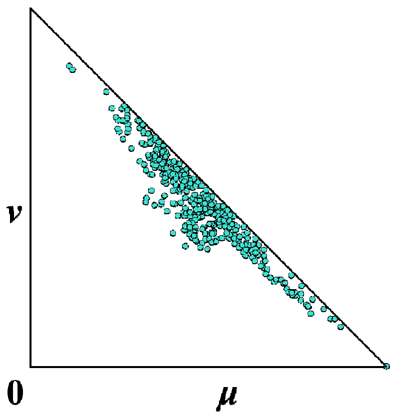


Figure 4: ICA over objects data with 10% errors

By taking a look at the results obtained by ICA over the original data set Fig. 1 it turns out that even 10% of errors in the data set does not significantly affect the method. It can be notices that the correlations are slightly reduced which can be observed in Fig. 3 as accumulation in the middle part of the intuitionistic triangle. This means that the intercrtieria correlations are slightly decremented – take for example criteria 5 and 12. In Fig. 4 the removal of consecutive objects leads also to reduction of correlations and appearance of a new relation between criteria 6 and 11. No significant incrementation of uncertainty level can be accounted in both cases. That confirms the high accuracy of ICA approach and makes it ideal in areas with missing or incorrect data.

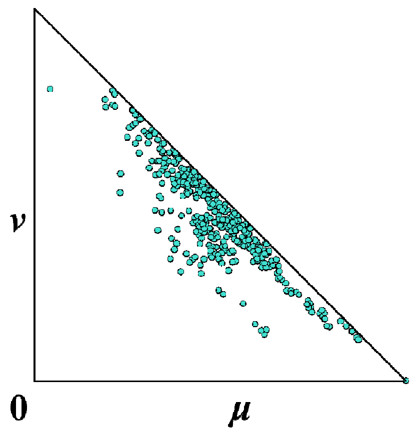


Figure 5. ICA for 29 consecutive missing objects

Fig. 5 illustrates the impact of 29 consecutive missing objects for criterion number 11. The level of uncertainty is increased as can be noticed in the figure. This number of missing criteria is equivalent of 87 hours long failure. This scenario is plausible for the remote measuring stations.

4 Conclusion

The conducted research showed that the intercriteria approach performs well in the presence of errors and missing data. No significant changes were observed in the case of error level lower than 1%. By increasing the error rate up to 10% a reduction in the intercriteria correlations was noticed as well as an increase of uncertainty Nevertheless it is always preferable to operate with correct readings and avoid using data sets with incomplete rows and pillars. The results confirm the ICA effectiveness over the proposed data set. More tests over variety of data are required in order to confirm the posted results.

Acknowledgements

This paper is partially supported by Grant DFNI-I-02-5/2014 of the Bulgarian National Science Fund, and by Project NIH-355/2015 of Prof. Asen Zlatarov University, Burgas. The measurements are kindly provided by [4, 5].

References

[1] Atanassov, K., D. Mavrov, V. Atanassova. Intercriteria Decision Making: A New Approach for Multicriteria Decision Making, Based on Index Matrices and Intuitionistic Fuzzy Sets. Issues in Intuitionistic Fuzzy Sets and Generalized Nets, Vol. 11, 2014, 1–8.

[2] Atanassova, V. Interpretation in the Intuitionistic Fuzzy Triangle of the Results, Obtained by the InterCriteria Analysis. 16th World Congress of the International Fuzzy Systems Association (IFSA), 9th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT), 30.06-03.07.2015, Gijon, Spain, pp. 1369–1374, doi:10.2991/ifsa-eusflat-15.2015.193.

[3] Kostadinov, T., T. Petkov An example of intercriteria analysis application to weather parameters. Notes on Intuitionistic Fuzzy Sets, Vol. 21, 2015, No. 2, pp. 126–133, ISSN 1310–4926.

[4] http://www.stringmeteo.com/synop/bg_stday.php

[5] http://www.burgas.bg/bg/info/index/591.