5th International Congress on Scientific Research April 21-22, 2024, Türkiye by IKSAD Institute

Homogeneity analysis for sustainable regions – case study: Mezőföld

Dániel Váradi (https://orcid.org/0000-0001-9610-8566),
László Pitlik (Jr.) (https://orcid.org/0000-0002-8058-9577),
Mátyás Pitlik (https://orcid.org/0000-0002-1991-3008),
Dr. László Pitlik (https://orcid.org/0000-0001-5819-0319)
e-Mails: danielvaradi140@gmail.com,, ptlklszl@my-x.hu, ppk@my-x.hu, pitlik@my-x.hu
Kodolányi János University and MY-X research team Hungary
Keywords: AI, anti-discrimination, optimization, automation, similarity analysis, group building

Introduction

History: The objective evaluation, the AI-based term-creatin in case of abstract phenomena is a relevant part of the research activities of the MY-X research team (c.f. https://miau.my-x.hu/miau2009/index.php3?x=e0&string=homogeneity.of.c). The GPS (general problem solver-oriented) AI does need the capability of the handling the terms of "Good<Better<Best" and also the derivation of arbitrary abstractions (terms) based on measurable variables. The human experts talk about relatively homogeneous/sustainable regions since ever (c.f. Mezőföld). The same question is interpretable for groups of human individuals (e.g. teams in sport and/or in enterprises/institutions). The subjective feelings about group/team-homogeneity should be supported through AI-solutions, where (here and now) groups of settlements will be analysed in order to derive which settlement-constellation can be defined as the most sustainable group – let alone in time-series-view.

Background and benchmarks: The observed region (Mezőföld) has 37 settlements having statistical data for 4 decades (1992-2002-2012-2022). 15 settlements can be interpreted as a kind of determining settlement concerning the form of the mapped polygon of Mezőföld. The Hungarian Statistical Office (KSH) has an online statistical service called TEIR. The regional statistics deliver data for relatively few phenomena concerning 4 decades: e.g. Housing stock (pcs) Children enrolled in kindergarten (person) Places to perform tasks in kindergarten (person) Kindergarten places (person) Groups of children in kindergarten (person) Kindergarten teachers (person) Divorces (cases) Domestic emigration (permanent and temporary together) (person) Domestic immigration (permanent and temporary together) (person) Infant mortality (died under 1 year) (person) Live births (person) Deaths (person) Marriages (case).

Highlighted details: The mathematical challenge is trivial: How the arbitrary dimensions (see 13 statistical variables/descriptors) can be aggregated to a kind of homogeneity index in an optimized way? The antidiscriminative optimization makes possible to derive the expected index values for different objects. Objects are the entire group of the 37 settlements (O1) and each constellation (O2-O16) where always one (i) of the 15 edge-settlements will be excluded from the calculations of the standard deviations (see group36_i). The lower is the standard deviation concerning each of the 13 attributes the higher is the aggregated homogeneity. The hypothesis is: can we evaluate each object with the same homogeneity index or not?

The results are interesting: The homogeneity index for the entire group of the 37 settlements is the highest one. Each reduced group is less sustainable. This is acceptable because the standard deviation of smaller or wider groups can be arbitrary high/low compared to each other. The results are: the most sensitive settlements are the middle voluminous ones (like Dunaföldvár > Enying > Tamási based on the average risks and Paks > Dunaújváros based on the higher standard deviations concerning the time-series values. Parallel: the most stable year is 1992. The less stable year is 2012 (based both on risk averages and standard deviations). The most sensitive settlement is Fadd (1992), Dunaföldvár/Paks (2002), Enying (2012, 2022). All these results belong to the calculation where the standard deviation for the group of 37 settlement was modified (36/37%) compared to the groups with 36 settlements.

Future aspects: After closing the manual-driven test-cases, the entire evaluation process can be automated e.g. in frame of development task for a bachelor's degree.

Literature, backgrounds

Already the own research activities in the last decades delivered useful impulses for the question: what can be seen as a sustainable region (group of settlements)? The first impulse is the mathematical interpretation of the term "sustainability". This term is mostly used in political contexts, but these interpretations are subjective and fuzzy. Sustainability must be and is a term with mathematical background. Sustainable is a system in a static view, if each available data about it can be derived from the other available data: c.f.

- https://miau.my-x.hu/miau2009/index.php3?x=e0&string=of.sustainability (mathematics of the sustainability)
- https://miau.my-x.hu/miau2009/index.php3?x=e0&string=szocio- (socio-physics)
- https://miau.my-x.hu/miau2009/index.php3?x=e0&string=aesthetic (mathematic of the beauty)

In a dynamic view, sustainability is a process, where the direction of the changes can be estimated based on the differences between the measured reality and the expected norm values: e.g. the price/performance analyses can detect price-advantages and price-disadvantages, and also norm-like prices. The disadvantageous prices will be reduced, the advantageous ones will be increased in long-term…

The second impulse is the similarity analysis as such. This AI-methodology (c.f. Q-GPS = quasi general problem solving) is capable of producing models (expert systems, simulators, production functions, neural networks, etc.) based on staircase functions for arbitrary contexts or even in a context free way. The similarities between objects based on their attributes make possible to estimate each given data position through the other (remained) data. This leads to a parallel data universe, where more or less differences can be detected between the measured universe and the estimated universe. The similarity analysis makes possible to handle the term of (acoustic and/or visual) beauty and/or the physics-like interpretation of seemingly more complex e.g. historical/sociological processes.

The third (here and now the last) impulse from the past of the research activities is the automation of the SWOT-technique. The SWOT analyses are in general more than subjective: quasi each idea can be set to each character (S_W-O-T) with a little intelligence for argumentation… This is a kind of chaos and not a scientific approach. BUT: the SWOT-analysis can be automated.

The steps of the analysis

This chapter presents following sub-chapters:

- Map
- OAM (object-attribute-matrix) parameters
- Process: Steps of the homogeneity analyses

The map

Figure#1: The observed region with 37 settlement (and 15 polygon-building-settlements) – Source: own presentation

OAM

The case study works with the following parameters:

Objects

Number of the objects = 64=4*[1+15] rows in the OAM, where 16 = 1 * group_37 & 15 * group_36 (excl. 1 single polygon-settlement), and 4 = 4 decades (1992, 2002, 2012, 2022)

Attributes

Source: KSH-TEIR (Központi Statisztikai Hivatal / Statistical Central Office, Regional Information System) - https://www.oeny.hu/oeny/teir/#/, where 13 variables (columns in the OAM) are available for all time periods =1992-2002-2012-2022 (4 decades). All attributes are compared to the population of the settlements (units = original unit of the focused attribute / capita).

List and units of the attributes:

1. Housing stock (pcs)
2. Children enrolled in kindergarten (person)
3. Places to perform tasks in kindergarten (person)
4. Places in kindergarten (person)
5. Kindergarten groups (person)
6. Kindergarten teachers (person)
7. Divorces (cases)
8. Domestic emigration (permanent and temporary together) (person)
9. Domestic immigration (permanent and temporary together) (person)
10. Infant mortality (died under 1 year) (person)
11. Live births (person)
12. Deaths (person)
13. Marriages (case)

Steps of the homogeneity analyses

- Raw OAM (see before: 64 rows * [13+1] attributes)
- Relativised OAM (unit for each raw data / capita based on the population of the settlements in the given decade)
- Standard deviation of the relativised attributes based on the given object (settlement-group) with different number of settlements (see before: 37 settlements * 1 case (all), 36 settlements * 15 cases)
- Ranked OAM (the less st.dev, the more sustainability/homogeneity – ranking always within the 64 objects)
- COCO Y0 (optimized online antidiscrimination models - https://miau.my-x.hu/myx-free/)
- Validation (based on symmetry of the direct & inverse staircases)
- Hermeneutics / Data – visualisation

More details for the entire reproducibility: https://miau.my-x.hu/miau/311/mezofold/

Results

This chapter is the most interesting challenge concerning the expected/offered automation. The challenge is: which kind of (context-free) interpretation rules can be identified in order to enforce appropriate text schemes.

valid	1					
Units: index values						
estimations	**decades**					
objects	**1992**	**2002**	**2012**	**2022**	**average**	st.dev
only_36_Adony	1000038	999963	999959	999971	999983	37
only_36_Bölcske	999994	1000038	999948	999955	999983	41
only_36_Dunaföldvár	999985	999950	999966	999962	999966	14
only_36_Dunaszentgyörgy	999993	999961	999964	999959	999969	16
only_36_Dunaújváros	1000038	1000060	999960	1000080	1000034	52
only_36_Enying	999977	999984	999948	999954	999966	17
only_36_Fadd	999967	999976	999976	1000037	999989	32
only_36_Madocsa	999989	999963	1000038	999961	999988	36
only_36_Mezőkomárom	1000038	1000038	1000038	1000037	1000038	0
only_36_Paks	1000038	999950	999964	1000063	1000004	55
only_36_Pusztaszabolcs	999973	999967	999977	999963	999970	6
only_36_Szedres	1000038	1000038	1000038	999967	1000020	35
only_36_Székesfehérvár	999979	999968	999972	999999	999979	14
only_36_Tamási *	999979	999957	999951	999988	999969	17
only_36_Tengelic	999989	999998	999980	999983	999987	8
all_37	1000121	1000162	1000196	1000141	1000155	32
average	**1000008**	**999998**	**999992**	**1000001**	**1000000**	<--norm
st.dev.	41	57	63	55	47	

Figure#2: Estimated homogeneity index values – Source: own presentation

Figure#2 presents the estimated homogeneity index values, their averages for columns (decades) and rows (objects), and the standard deviations too.

The hermeneutical sub-system needs rules: e.g.

- IF the most green cells in the columns can be seen in the same row AND this row is even the scenario (object), where all settlements are involved, THEN the conclusion is that the most sustainable constellation needs all the settlements. This direct conclusion has further indirect implications: e.g. the region can be revolvered – especially through the most sensitive settlements (c.f. weakest members).
- (Further rules are necessary to cover all the possibilities concerning the above-mentioned combinatorial space about the most green cells and their interpretations.)
- IF the average score of a settlement (columns) is massive red AND below the norm value of 1000000, THEN this/these settlement(s) are the most sensitive ones. It means that the exclusion of these settlements can cause the most instability for the region.
- IF the average score of the rows (objects) is massive green, THEN this decade can be interpreted as the most stable period and vice versa: the most red decade is the most instable period.
- IF the standard deviation of the rows (objects) is massive green, THEN this decade can be interpreted as the most stable period and vice versa: the most red decade is the most instable period.
- IF the standard deviation of the columns (decades) is massive green, THEN this object (part of the region) can be interpreted as the most irrelevant constellation and vice versa: the most red object is the most sensitive one – with blackmail-potential...
- IF the most red objects (in different columns and/or in the aggregated views like average and/or standard deviation) are quasi rel. little town (like Dunaújváros, Paks, Dunaföldvár, Enying), THEN this type can be seen as the most critical components in the building process of regions (big cities and small villages are irrelevant concerning the regional sustainability).
- …
- IF the validity of the estimation is for all objects are given, THEN the raw data and the methodology can be interpreted as qualitative matured.

SWOT	1992	2002	2012	2022	trend
only_36_Adony	S	W	W	W	T
only_36_Bölcske	W	S	W	W	T
only_36_Dunaföldvár	W	W	W	W	T
only_36_Dunaszentgyörgy	W	W	W	W	T
only_36_Dunaújváros	S	S	W	S	O
only_36_Enying	W	W	W	W	T
only_36_Fadd	W	W	W	S	O
only_36_Madocsa	W	W	S	W	T
only_36_Mezőkomárom	S	S	S	S	T
only_36_Paks	S	W	W	S	O
only_36_Pusztaszabolcs	W	W	W	W	T
only_36_Szedres	S	S	S	W	T
only_36_Székesfehérvár	W	W	W	W	O
only_36_Tamási *	W	W	W	W	O
only_36_Tengelic	W	W	W	W	T
all_37	S	S	S	S	O

Figure#3: SWOT-view of the estimated homogeneity index values – Source: own presentation

Figure#3 presents the S_W_O_T-view of the estimated homogeneity index values (incl. the trend-based interpretation concerning Opportunity and/or Threatening).

The hermeneutical sub-system also needs rules: e.g.

- IF a potential regional object has the pattern WWWWT, THEN the excluded settlement is dangerous concerning the sustainability of the entire region.
- IF the object with all the settlements has the pattern SSSSO, THEN the sustainability for the entire region is high and the trend is increasing.
- (The whole combinatorial space for the S_W_O_T letters must have rules without any inconsistency.)
- (The all_37 constellation should not be interpreted in frame of the SWOT-analysis, because this benchmark region needs other interpretation text schemes as the other ones [36_*]).

Discussion

This paper presented a small focus about the grouping settlements (building groups), because one single region has been analysed and estimated whether one of the elements should be excluded from the group and which elements have a kind of sensitivity. It means, which elements (settlements) can cause more damage through their exclusion than in cases of other ones?

There are however more complex views: e.g. the view of the more-region-modelling-approach, where the goal is to build more region from a quasi unlimited set of elements. In this process, it is possible that one single settlement do build a region (c.f. city-states like Athen).

Conclusion

The optimized building of groups is a kind of clustering, but the mathematical background is an other one, than before. The staircase functions make namely possible to derive validities of object and/or model-levels. Validity means: whether we might interpret the results based on the given data or in case of some objects we can not see the expected consistency between the different model-layers. Staircase functions produce symmetry-effects, and in case of a direct and an inverse (mirrored) input data set, the result should also be mirrored if the raw data are qualitative and quantitative good enough.

Future

The derivation of the optimized sensitivity of the system-elements (settlements) will need further modelling steps, where new attributes of the already analysed objects can be defined (see trend, max, min, standard deviation, average, etc.) in order to derive a multi-dimensional and anti-discriminative optimum based again on the similarity analysis (COCO Y0).

The entire introduced analytical steps can be automated and therefore can be used as a general tool for decision support (on the field of e.g. data-driven policy making, HR/military/sport (team-building), etc.

Parallel, it could also be analysed which external settlement(s) (see as 38th element/member) would be worth including to the benchmark region (all_37)?

References

…all references can be found in the text…