Distance Metrics Library for MCDA Methods

Aleksandra Bączkiewicz
University of Szczecin
Szczecin, Poland *aleksandra.baczkiewicz@phd.usz.edu.pl*

Jarosław Wątróbski
University of Szczecin
Szczecin, Poland *jaroslaw.watrobski@usz.edu.pl*

Wojciech Sałabun
National Institute of Telecommunications
Warsaw, Poland *w.salabun@il-pib.pl*

Abstract

Information systems based on Multi-Criteria Decision Analysis (MCDA) methods enable considering multiple attributes with contrary objectives. Information systems using MCDA simplify and automatize assessment toward automatizing decision support systems. Individual MCDA methods differ in their algorithms, implying different results for the same problem. Moreover, the diversity of algorithms refers to the MCDA methods and their techniques used at an individual stage, such as distance metrics. They are implemented in MCDA methods to measure alternatives' distances from reference solutions. The most commonly used metric is the Euclidean distance. However, other distance metrics are also suitable for this purpose. Moreover, a broad set of metrics can be helpful in comparative analysis to test the robustness of particular scenarios. Therefore, the main contribution of a Python library for multi-criteria decision analysis called distance-metrics-mcda is providing a set of 20 distance metrics for benchmarking purposes. The implemented library offers an autonomous tool for evaluating any decision problem. The presented library is an important addition to decision support systems based on MCDA methods as it provides additional possibilities for analysis of scenarios' reliability.

Keywords: Distance Metrics, MCDA, TOPSIS, Decision Support Systems

1. Introduction

Modern information systems are primarily based on scientific methodologies, such as machine learning, big data, data mining, and Multi-Criteria Decision Analysis (MCDA) methods. In the MCDA domain, selecting the most suitable method for the problem under solution is widely discussed. Besides, methods with multiple algorithm versions may provide different results for the same decision problem. For this reason, decision-makers and researchers often use comparative analysis, considering various methods giving multiple scenarios. Furthermore, receiving similar results of a given scenario using different methods confirms its robustness and stability in the presence of possible changes introduced to the model. On the other hand, achieving outstanding results suggests the necessity of performing additional analytical procedures. In this situation, various distance metrics for benchmarking provide a valuable supplement to the well-known and widely used Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method, which has proven its effectiveness in decision-making problems in multiple domains [20].

This paper presents a Python library providing methods for multi-criteria decision analysis focusing on distance metrics that measure the distance between compared solutions containing crisp numerical values. For MCDA methods such as TOPSIS or COmbinative Distance-based ASsessment (CODAS), based on measuring the distance from reference solutions to evaluate alternatives, distance metrics are used to determine sought distance. For example, the TOPSIS method uses the distance measurement of alternatives from Positive and Negative Ideal Solution [12]. In contrast, the CODAS method measures the distance of solutions from the Negative Ideal Solution [11]. This paper aims to demonstrate the usage of distance metrics other than the most commonly used Euclidean distance to measure the distance to reference solutions in the MCDA domain. In MCDA methods, the most widely used distance metric is Euclidean distance, so investigating the effect of different metrics used for the same purpose is an interesting research gap. Another goal of this work includes demonstrating the application of different metrics to examine the robustness of solutions. The application of different distance metrics in the MCDA domain was explored using an illustrative case of mobile phones.

The Euclidean, Manhattan, or Chebyshev distances are most commonly applied. It is observed that Manhattan and Euclidean metrics proved to give the most consistent results and appear to be the most appropriate for TOPSIS. However, most studies have not considered other distance metrics available in the literature, which are also suitable for use in the TOPSIS algorithm [19]. Because of this, access to a set containing more than a few distance metrics gives a decision-maker who uses TOPSIS to evaluate a multi-criteria problem more possibilities [17]. Authors of mentioned works among metrics useful in distance-based MCDA methods utilizing crisp performance values propose set including Manhattan, Euclidean, Chebyshev, Squared Euclidean, Sørensen or Bray-Curtis, Canberra, Lorentzian, Jaccard, Dice, Bhattacharyya, Hellinger, Matusita, Squared-chord, Pearson χ^2 and Squared χ^2 distance metrics for using with the TOPSIS method. Supplying multiple distance metrics allows for evaluating a given decision's stability and robustness. Using various distance metrics to measure the distance of alternatives from reference solutions may result in different solutions. Various distance metrics use different formulas to calculate the distance between compared solutions. If the result does not change significantly despite multiple distance metrics, they suggest high stability and robustness to changes in a given alternative. A decision support system providing multiple metrics allows decision-makers and researchers to experiment and conduct simulations. The library implemented by the authors has the potential to serve as an engine for the decision support system. Furthermore, the broad set of implemented distance metrics used by distance measuring based MCDA methods like TOPSIS or CODAS gives the possibility of complementing the decision procedure with an extended analysis of the reliability of the considered solution, which helps confirm the best option. Due to their wide range, the metrics implemented in the developed library can be used not only in the field of MCDA but also in other areas, such as artificial intelligence, to determine the distance between solutions and which of them is closest to the sought solution.

The rest of the paper is organized as follows. Section 2 provides a literature review on the topic under research. Next, the applied methodology is given in section 3. Then, in section 4 obtained results are presented and discussed. Finally, in the last section 5 conclusions and a summary are given, and future work directions are drawn.

2. Literature Review

Many distance metrics utilizing different formulas can be found in the literature. They are used in measurements in various fields. Euclidean distance is one of the best-known and widely used distance metrics in MCDA methods based on measuring the distance to reference solutions. This measure is used as an implicit metric to calculate the distance of the options considered from the positive ideal (PIS) and negative-ideal (anti-ideal, NIS) solutions in the TOPSIS method algorithm [12]. The Euclidean distance is also the principal metric for measuring the distance of individual options from the anti-ideal solution in the case of the CODAS method [13].

The Euclidean distance metric has evolved to several modifications of the distance calculation formula, including Square Euclidean distance and Standardized Euclidean distance. These metrics find implementation in multi-criteria decision-making areas such as data selection, criteria significance determination, and establishing variability among personalized and ingroup preferences [22]. Like Euclidean distance, Cosine distance is a widely used metric in vector models. Research benchmarking Euclidean distance and Cosine distance in various engineering areas, for example, image recognition process, demonstrates comparable effects provided by investigated metrics. Moreover, cosine distance is applied in many domains where features comparison is required, for example, for image processing or voice recognition [9]. Among popular distance metrics, the correlation distance can be mentioned. This measure is used in investigations involving data for most relevant feature selection and comparing numerically expressed objects, such as vectors [6]. Exploring distance metrics exposes the Chebyshev distance, which is alternatively called chessboard distance or maximum metric. Chebyshev distance determines the distance between two vectors by calculating the maximum absolute distance along any coordinate dimension. This measure is incorporated into practical problems such as chess, classification, warehouse logistics, signal processing, intelligent recognition applications, and many others [4]. Mentioned papers demonstrate that distance metrics are applicable not only in the MCDA domain but also in computer vision. This field employs Hausdorff distance, which is suitable for pattern recognition, more precisely to establish the similarity degree between compared images [21]. Further usage of Hausdorff distance compares vehicle trajectories represented by two sets of points. The indicated method is applied in vehicle trajectory analysis utilizing computer vision techniques. This methodology is practically employed to detect potentially hazardous traffic behaviours [5]. Distance metrics are employed in the development of well-known MCDA method extensions. For example, the Chebyshev metric has been applied for Pythagorean membership degrees in the ELECTRE method to solve multi-criteria decision problems under Pythagorean fuzziness [4]. Extending existing methods with new ideas requires a comprehensive benchmarking analysis to assess their reliability. Exploring the usability of alternative distance metrics in MCDA algorithms is justified by limitations, including a lack of taking into account correlation between criteria by Euclidean distance in the TOPSIS method, which may result in distorted outcomes due to overlapping information. Besides, the lack of distance weights incorporation is also criticized. It became the motivation for the research considering alternative metrics for example weighted Euclidean distance [1], Euler distance [3], newly developed distance measures involving entropy and difference coefficients [18] and many other approaches [3]. Noteworthy work focused on replacing Euclidean distance in the TOPSIS algorithm is the topic of paper [14]. In this research, Euclidean distance was replaced by a grey correlation coefficient. This modification was performed to grasp potential uncertainty more accurately. The authors of paper [17] suggest the need to provide a tool that offers the possibility of different metrics in the TOPSIS method for decision-makers. Such a tool allows conducting a comparative analysis of scenarios and confirming their reliability. Furthermore, the tool containing a wide range of distance metrics for the TOPSIS method can play an educational role for people studying MCDA methods and serve practitioners in solving decision-making problems.

The explored literature proves that selecting a methodology for multi-criteria problem assessment is a complex procedure. Moreover, differences between assumptions and formulas of particular methods and their effect on the final results make it impossible to indicate better or worse approaches. Therefore, an appropriate choice of methodology usually requires additional preliminary experiments to limit the number of methods to consider and evaluate the validity and reliability of best-scored scenarios.

3. Methodology

The implemented Python library named ‘distance-metrics-mcda‘ provides 20 distance metrics and other methods necessary for multi-criteria decision making. The developed library includes the TOPSIS method, which measures the distance of alternatives from reference solutions. The

TOPSIS method utilizes distance metrics selected from the provided set in this aim. Additionally, five decision matrix normalization techniques (Linear, Minimum-Maximum, Maximum, Sum, and Vector normalizations method), three correlation coefficients to determine the similarity of the obtained solutions (Spearman, Weighted Spearman Rank Correlation Coefficient, and Pearson correlation coefficient). The library also contains two popular objective methods for determining criteria weights, the Entropy weighting method [15] and the Criteria Importance Through Inter criteria Correlation (CRITIC) weighting method [2]. These methods may be used instead of subjective weighting methods requiring the involvement of the decision-maker. Some subjective weighting techniques have a complex algorithm, such as AHP, and the decision-maker does not always have sufficient expert knowledge. In such situations, objective weighting techniques that determine the weights of the criteria based on the data in the decision matrix using mathematical formulas prove to be useful. By including objective weighting methods in the implemented library, the lack of expert knowledge of the decision-maker will not be a limitation to its usage. Besides, the objective weighting techniques make this tool autonomous, which is one of the authors' main goals in contribution to the developed library. Figure 1 presents a framework demonstrating the flowchart performance of using this library.

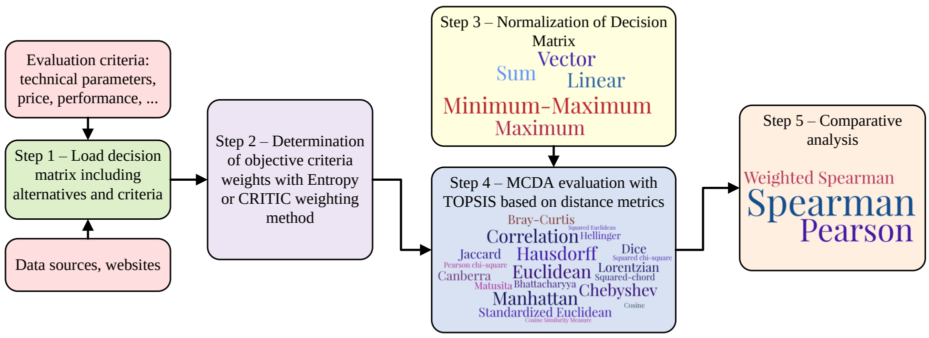

Fig. 1. Flowchart presenting framework of using the implemented library.

The developed library can be downloaded from [7] and installed using the pip command. Complete codes and usage examples are available on GitHub at [8]. The mentioned repository also includes examples and sample code helpful in visualizing the results, enabling graphs like the one in this article. The basic assumptions and formulas of methods provided in presented library and applied in this research are provided on GitHub in Supplementary Material.

The TOPSIS algorithm measures the geometric distance of alternatives from an ideal and an anti-ideal reference solution. The method assumes that the best option is closest to the ideal solution and farthest from the anti-ideal solution, and its algorithm is provided in [12]. TOPSIS, like most other MCDA methods, requires normalization of the decision matrix and providing the types of criteria and their weights. The TOPSIS method returns the scores of the alternatives, namely preference function values. The best-ranked alternative has the highest preference function value. Therefore, the alternatives ranked by the TOPSIS method are sorted by preference value in descending order.

4. Results

As an example of applying the 'distance-metrics-mcda' library, the multi-criteria problem of selecting the most advantageous mobile phone model is presented. The data of the evaluated models considering 11 selection criteria were acquired from the reference paper [10]. The released collection contains data on the performances and technical parameters of 25 mobile phones. To illustrate the performance and usefulness of the developed library explicitly, the authors selected the first 15 models of mobile phones. The alternatives' evaluation criteria and performance values are presented in detail on GitHub. The evaluation criteria, which are the technical attributes of the assessed models, are provided in Table 1.

Table 1. Evaluation criteria of mobile phones considered in this research.

G	Criteria group	C_j	Explanation	Type
G_1	Hardware and performance	C_1	Front camera resolution (megapixels)	Max
		C_2	Rear camera resolution (megapixels)	Max
		C_3	Battery capacity (mAh)	Max
		C_4	RAM (GB)	Max
		C_5	Screen size (inch)	Max
		C_6	CPU rating	Max
G_2	Appearance	C_7	Appearance rating	Max
G_3	Brand	C_8	Market share (%)	Max
		C_9	Brand favorable rate (%)	Max
G_4	Accessory	C_{10}	Accessory rating	Max
G_5	Price	C_{11}	Price (RMB)	Min

Type Max represents profit criteria with a maximization aim. On the other hand, Min denotes the minimalization aim for cost criteria. The decision matrix fragment containing evaluated mobile phones' performance values is displayed in Table 2. In addition, the complete dataset is provided in the GitHub repository.

Table 2. Fragment of the decision matrix with alternatives performance values.

A_i	Name	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}
A_1	Huawei Honor V10	13	2	3750	6	6	6701	3.2	9.8	0.72	2.9	2999
A_2	Samsung Galaxy Note8	8	12	3300	6	6.3	6806	4.3	12.7	0.82	3.7	6988
A_3	iPhone8 Plus	7	12	2675	3	5.5	10304	3.4	7.8	0.86	3	6688
...	...	...	...	...	...	...	...	...	...	...	...	...

The decision matrix containing the alternatives' data was normalized using the Minimum-Maximum method implemented in the function called 'minmax_normalization', and the CRITIC objective weighting method implemented in the function called 'critic_weighting' was used to determine the weights. The alternatives were then evaluated with TOPSIS using ten distance metrics selected from developed library: Euclidean, Manhattan, Bray-Curtis, Canberra, Lorentzian, Hellinger, Matusita, Squared chord, Pearson chi-square, and Squared chi-square distance metrics. Multi-criteria assessment of alternatives was performed with the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method, using criteria weights calculated with a chosen weighting method. The box chart in Figure 2 visualizes the distribution of preference function values calculated for evaluated alternatives by the TOPSIS method using ten selected distance metrics. It is worth noting that the A_8 (Oppo R11s) mobile phone received the highest preference values. High preference values were also observed for the A_9 (Huawei Mate10 Pro-) and A_2 (Samsung Galaxy Note8). Figure 3 compares the rankings of the evaluated alternatives obtained with the TOPSIS method using ten selected distance metrics.

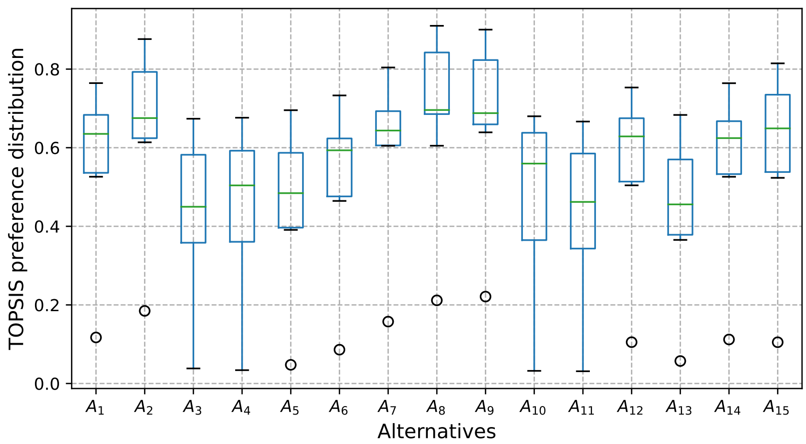

Fig. 2. TOPSIS preference values determined with different selected distance metrics.

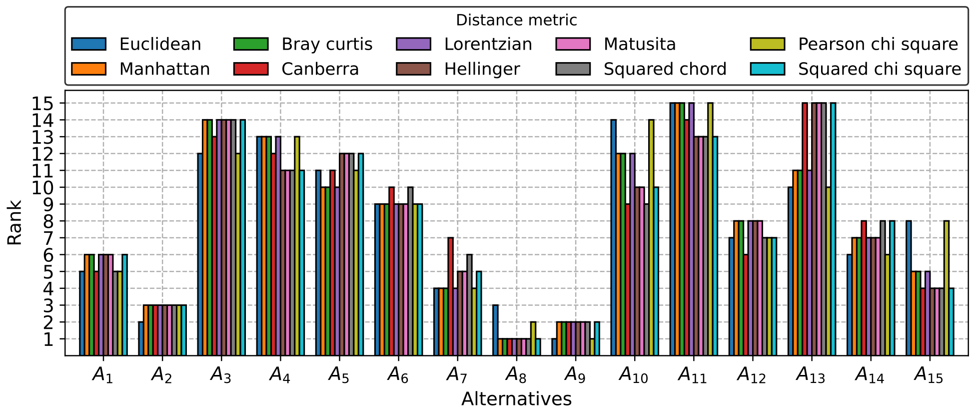

Fig. 3. Comparison of rankings generated by TOPSIS using different distance metrics.

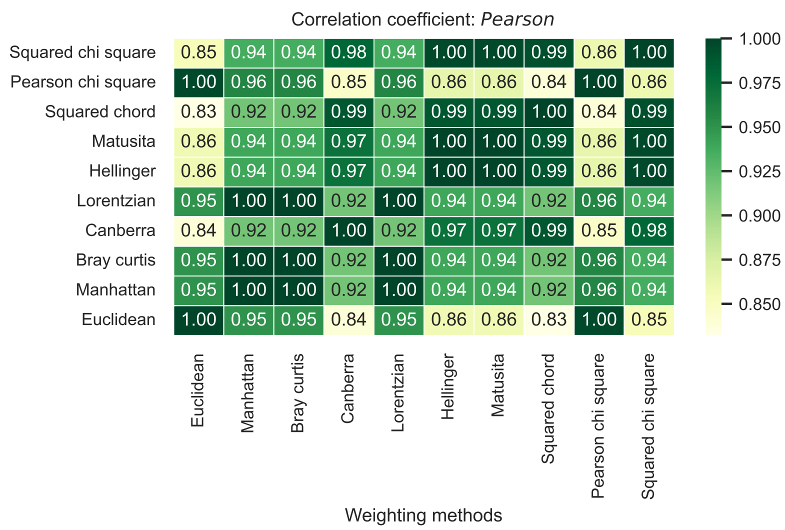

Fig. 4. Correlations of rankings determined by TOPSIS using different distance metrics.

Results are consistent with outcomes in the box chart. The highest ranks were achieved by A_8, A_9, and A_2, respectively. The results confirm the stability of the alternative A_8, which was the ranking leader in the case of simulations with eight metrics out of 10 selected. It can be observed that, in fact, often used Euclidean distance did not rank A8 as the best alternative. This observation justifies the relevance of relying on more than one popular distance metric. Furthermore, it demonstrates the robustness of this solution to potential changes that may occur in the model due to small changes in the values of the weights or inaccuracies in the data.

The final step of the multi-criteria analysis of the exemplary dataset with mobile phones was to determine the correlation values between the rankings of phones evaluated by the TOPSIS method using different distance metrics. The correlation results were calculated using Pearson's correlation coefficient. The correlation values visualized in Figure 4 are high, indicating high convergence of the results obtained using different distance metrics. The high consistency confirms the robustness of the evaluation model.

5. Conclusion

The presented experimental results carried out using the 'distance-metrics-mcda' library confirm the usefulness of this tool in solving multi-criteria problems where the decision-maker needs to determine the robustness of the solutions. The main contribution of the presented library is 20 different distance metrics, which are an important part of the algorithm of the well-known and widely used TOPSIS method. However, applying the set of provided distance metrics is not limited to the TOPSIS method. The distance metrics are used analogously by another multi-criteria decision analysis method, CODAS, which measures the distance of the alternatives from the anti-ideal solution, and by the NAIADE method [16]. This fact indicates directions for further work to expand the library to include other methods using distance metrics, enabling decision-makers and researchers to perform more simulations that also consider multi-criteria methods other than TOPSIS during decision analysis. The distance metrics can also be useful in stochastic algorithms, such as evolutionary, genetic, and swarm algorithms, helpful in seeking the best solutions in many domains. They can provide a tool to measure the distance of solutions in the goal function of stochastic algorithms, indicating that the distance metrics implemented in the library have broader applications than the MCDA domain.

Acknowledgements

The work was supported by the project financed within the framework of the program of the Minister of Science and Higher Education under the name "Regional Excellence Initiative" in the years 2019-2022, Project Number 001/RID/2018/19; the amount of financing: PLN 10.684.000,00 (A.B. and J.W.).

References

1. Arslan, T.: A weighted Euclidean distance based TOPSIS method for modeling public subjective judgments. Asia-Pacific Journal of Operational Research. 34(03), pp. 1750004 (2017)
2. Bączkiewicz, A., Kizielewicz, B., Shekhovtsov, A., Wątróbski, J., Sałabun, W.: Methodical aspects of MCDM based E-commerce recommender system. Journal of Theoretical and Applied Electronic Commerce Research 16(6), pp. 2192–2229 (2021)
3. Çelikbilek, Y., Tüysüz, F.: An in-depth review of theory of the TOPSIS method: An experimental analysis. Journal of Management Analytics. 7(2), pp. 281–300 (2020)
4. Chen, T.Y.: New Chebyshev distance measures for Pythagorean fuzzy sets with applications to multiple criteria decision analysis using an extended ELECTRE approach. Expert Systems with Applications. 147, pp. 113164 (2020)

5. Chen, Y., He, F., Wu, Y., Hou, N.: A local start search algorithm to compute exact Hausdorff Distance for arbitrary point sets. Pattern Recognition. 67, pp. 139–148 (2017)
6. Edelmann, D., Móri, T.F., Székely, G.J.: On relationships between the Pearson and the distance correlation coefficients. Statistics & Probability Letters. 169, pp. 108960 (2021)
7. energyinpython: Python 3 library for Multi-Criteria Decision Analysis based on Distance Metrics (2022), `https://pypi.org/project/distance-metrics-mcda/`
8. energyinpython: Python 3 library for Multi-Criteria Decision Analysis based on Distance Metrics (2022), `https://github.com/energyinpython/distance-metrics-for-mcda`
9. George, K.K., Kumar, C.S., Sivadas, S., Ramachandran, K., Panda, A.: Analysis of cosine distance features for speaker verification. Pattern Recognition Letters. 112, pp. 285–289 (2018)
10. Guo, M., Liao, X., Liu, J., Zhang, Q.: Consumer preference analysis: A data-driven multiple criteria approach integrating online information. Omega. 96, pp. 102074 (2020)
11. Ijadi Maghsoodi, A., Ijadi Maghsoodi, A., Poursoltan, P., Antucheviciene, J., Turskis, Z.: Dam construction material selection by implementing the integrated SWARA—CODAS approach with target-based attributes. Archives of Civil and Mechanical Engineering. 19(4), pp. 1194–1210 (2019)
12. Jain, V., Sangaiah, A.K., Sakhuja, S., Thoduka, N., Aggarwal, R.: Supplier selection using fuzzy AHP and TOPSIS: a case study in the Indian automotive industry. Neural computing and applications. 29(7), pp. 555–564 (2018)
13. Karagoz, S., Deveci, M., Simic, V., Aydin, N., Bolukbas, U.: A novel intuitionistic fuzzy MCDM-based CODAS approach for locating an authorized dismantling center: a case study of Istanbul. Waste Management & Research. 38(6), pp. 660–672 (2020)
14. Lo, H.W., Hsu, C.C., Huang, C.N., Liou, J.J.: An ITARA-TOPSIS Based Integrated Assessment Model to Identify Potential Product and System Risks. Mathematics. 9(3), pp. 239 (2021)
15. Lotfi, F.H., Fallahnejad, R.: Imprecise Shannon's entropy and multi attribute decision making. Entropy. 12(1), pp. 53–62 (2010)
16. Munda, G.: A conflict analysis approach for illuminating distributional issues in sustainability policy. European Journal of Operational Research. 194(1), pp. 307–322 (2009)
17. Ploskas, N., Papathanasiou, J.: A decision support system for multiple criteria alternative ranking using TOPSIS and VIKOR in fuzzy and nonfuzzy environments. Fuzzy Sets and Systems. 377, pp. 1–30 (2019)
18. dos Santos, B.M., Godoy, L.P., Campos, L.M.: Performance evaluation of green suppliers using entropy-TOPSIS-F. Journal of cleaner production. 207, pp. 498–509 (2019)
19. Wątróbski, J., Bączkiewicz, A., Sałabun, W.: pyrepo-mcda—reference objects based MCDA software package. SoftwareX 19, pp. 101107 (2022)
20. Wątróbski, J., Bączkiewicz, A., Ziemba, E., Sałabun, W.: Sustainable cities and communities assessment using the DARIA-TOPSIS method. Sustainable Cities and Society p. 103926 (2022)
21. Zhang, J., Pang, J., Yu, J., Wang, P.: An efficient assembly retrieval method based on Hausdorff distance. Robotics and Computer-Integrated Manufacturing. 51, pp. 103–111 (2018)
22. Zou, Y., Chen, W., Tong, M., Tao, S.: DEA Cross-Efficiency Aggregation with Deviation Degree Based on Standardized Euclidean Distance. Mathematical Problems in Engineering. 2021, pp. 1–10 (2021)