Site selection for food distribution using rough set approach and TOPSIS method

Changez Khan[a], Sajid Anwar[a,*], Shariq Bashir[b], Abdul Rauf[c] and Adnan Amin[a]
[a]*Department of Computer Science, Institute of Management Sciences, Peshawar, Pakistan*
[b]*Deparment of Computer Science, Muhammad Ali Jinnah University (MAJU), Islamabad, Pakistan*
[c]*College of Computer and Information Sciences Al-Imam Muhammad ibn Saud Islamic University (IMSIU),
Riyadh, Saudi Arabias*

Abstract. Suitable site selection for a specific purpose is a crucial activity, and of the greatest importance to a project manager. Several methods have been proposed by the research community for effective site selection, but all proposed methods incur high costs. This study explores the combination of a rough set theory approach (RSTA) with Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) for suitable site selection for food distribution. This method provides a set of rules to determine different sites, which ultimately can help management develop strategies for suitable site selection. A set of rules for suitable site selection are derived from information related to a practical case, Pakistan Red Crescent Society (PRCS), to demonstrate the prediction ability of RSTA. The results clearly demonstrate that the RSTA model can be a valuable tool for site identification. Rough set theory also assists management in making appropriate decisions based on their objectives while avoiding unnecessary costs. However, while RSTA provides rules to determine the best sites for food distribution, it does not pinpoint the best sites for food distribution. To be more precise and accurate, this work is extended to another multi-criteria decision-making technique solution: the TOPSIS method. By using this method, this study provides the best top priority site for food distribution of PRCS.

Keywords: Site selection, rough set theory, TOPSIS, multi-criteria decision making (MCDM)

1. Introduction

With the rapid changes in the contemporary world environment, it has now become mandatory to focus on new areas of emerging needs for people worldwide. World citizens are running out of food due to wars, natural disasters, economic crises, etc. All these factors have created a situation in which people from all over the world, through international organizations, attempt to help disadvantaged people so that they can fight hunger effectively. One such organization, the Pakistan Red Crescent Society (PRCS), has been helping deserving people in most parts of the country. Most of the organizational activities are performed by PRCS in rural areas and at various sites around the country. The primary challenge to PRCS is the selection of appropriate sites for food distribution to Internally Displaced Persons (IDPs) in rural areas of Pakistan.

Site selection is a multi-criteria decision making problem for which PRCS takes into account a number of factors, such as security of the area, community involvement, technical aspects, locations, authorities, and environmental considerations.

Considering these factors, this study explores the use of a novel approach by combining the Rough Set Theory Approach (RSTA) with the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)

*Corresponding author. Sajid Anwar, Department of Computer Science, Institute of Management Sciences, Peshawar, Pakistan. Tel.: +92 91 9217408, 9217451; Fax: +92 91 9217407, 5861026; E-mail: sajid.anwar@imsciences.edu.pk.

method to select the most suitable sites for food distribution based on the above-mentioned factors. It has been assumed that the combination of these tools will help management make appropriate decisions based on a multi-criteria decision environment to achieve their objectives, and to avoid unnecessary costs.

The remainder of the paper is organized as follows. Section 2 presents a literature review, while Section 3 discusses preliminary studies. The development methodology is presented in Section 4, which is followed by the evaluation of methods in Section 5. Section 6 presents results and discussion, while Section 7 describes the strengths and weaknesses of the proposed research method. A summary of this paper is presented in Section 8, with experimental conclusions.

2. Literature review

The literature review indicates that many studies have been previously conducted on the subject of site selection and multi-criteria decision making. Arabani et al. [16] suggested an effective way to determine the best location for dams. Neshaei, et al. [19] investigated site locations of buried concrete water reservoirs. Cruz et al. [3] suggested a method of shipment scheduling. Sun, et al. [20] conducted research on site selection of passenger stations in the valley-city (China). Jing [27] conducted research regarding site selection for a logistics park. Important research on site selection of an aviation spare distribution center for military aircraft has been conducted by Chen, et al. [5]. Furthermore, Jingfeng, et al. [2] performed site selection decision making for a distribution logistics park. Moreover, site selection criteria for coastal tourism facilities have been proposed by Lo [6]. Alternatively, Lindgren [4] evaluated the selection of sites for development caused by the improper sitting of residential complexes and industrial facilities. In another study, Lawson [24] suggested the selection of a suitable site for an aquaculture venture. Another interesting research on site selection was also contributed by Yasser, et al. [17]. Moreover, a new addition to the field of site selection has been made by Elhadary, et al. [21], by introducing the rough set classification and attribute reduction method to landfill site selection. Srafaraz, et al. [23] developed a new hybrid MCDM technique to choose the locations of shopping malls in Tehran. In continuation, Majid, et al. [16] proposed an integrated MCDM SWARA–WASPAS method to select one city out of 25 in which to implement future solar power plants.

3. Preliminaries

3.1. Rough set theory

Rough set theory was originally proposed by Z. Pawlak in the early 1980s. Rough set theory deals with the classification analysis of data tables. The primary goal of rough set analysis is to synthesize the approximation of concepts from acquired data [30].

More details about rough set theory can be found in work done by Walczak, et al., Komorowaski, et al., Yao, et al., K. Kaneiwa, Dey, et al., Rissino, et al., Zhang, et al., Yiyu, et al. and Pawlak [1, 8, 10, 12, 19, 22, 26, 28, 29].
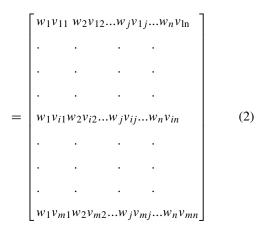
3.2. TOPSIS method

TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) was developed by Yoon et al. in 1980 [13, 14]. This method deals with multi-criteria decision making. The basic concept of the TOPSIS method asserts that the chosen alternative should demonstrate the shortest distance from the ideal solution and the farthest distance from the negative-ideal solution. The TOPSIS method is consists of the following steps:

Step 1: Construct the normalized decision matrix to transform the various attribute dimensions into non-dimensional attributes, which allows comparison across attributes.

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}} \tag{1}$$

Step 2: Construct the weighted normalized decision matrix.

$$v = \begin{bmatrix} v_{11} & v_{12}...v_{1j}...v_{\text{ln}} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ v_{i1} & v_{i2}...v_{ij}...v_{in} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ v_{m1} & v_{m2}...v_{mj}...v_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} w_1v_{11} \ w_2v_{12}...w_jv_{1j}...w_nv_{1n} \\ . \qquad . \qquad . \qquad . \\ . \qquad . \qquad . \qquad . \\ . \qquad . \qquad . \qquad . \\ w_1v_{i1}w_2v_{i2}...w_jv_{ij}...w_nv_{in} \\ . \qquad . \qquad . \qquad . \\ . \qquad . \qquad . \qquad . \\ . \qquad . \qquad . \qquad . \\ w_1v_{m1}w_2v_{m2}...w_jv_{mj}...w_nv_{mn} \end{bmatrix} \qquad (2)$$

Step 3: Determine the ideal and negative-ideal solutions.

$$A^+ = \{(\max_i v_{ij}/j \in J), (\min_i v_{ij}/j \in J) \big/$$

$$i = 1, 2, \dots m\} = \{v_1^+, v_2^+, \dots, v_n^+\}$$

$$A^- = \{(\min_i v_{ij}/j \in J), (\max_i v_{ij}/j \in J) \big/$$

$$i = 1, 2, \dots m\} = \{v_1^-, v_2^-, \dots, v_n^-\} \qquad (3)$$

where $J = \{j = 1, 2, \dots, n|$

j associated with benefit criteria}

$J = \{j = 1, 2, \dots, n|$

j associated with benefit criteria}

Step 4: Calculate the separation measure.
  – Ideal separation

$$S_i^+ = \sqrt{\sum_{j=1}^{n} (v_{ij} - v_j^-)^2} \ i = 1, 2, \dots\dots, m \qquad (4)$$

  – Negative ideal separation

$$S_i^- = \sqrt{\sum_{j=1}^{n} (v_{ij} - v_j^-)^2} \ i = 1, 2, \dots\dots, m \qquad (5)$$

Step 5: Calculate the relative closeness to the ideal solution.

$$C_i^* = \frac{S_i^-}{(S_i^+ + S_i^-)}, 0 < C_i^+ < 1, i = 1, 2, \dots, m$$

$$C_i^* = 1 \ if \ A_i = A^+ \qquad (6)$$

$$C_i^* = 0 \ if \ A_i = A^-$$

Step 6: Rank the preference order. A set of alternatives can now be preferentially ranked according to the descending order of C_i^*.

4. Development methodology

In this study, four distinct districts from Khyber Pakhtunkhwa (KPK) are selected for site selection by PRCS: Hangu, Nowshehra, Lower Dir and Charsadda. Hangu and Lower Dir share disaster characteristics, as do Nowshehra and Charsadda. Table 1 depicts conditional attributes, classifications of conditional attributes according to corresponding factors and corresponding decision levels in which H represents High, M represents Medium, L represents Low and N represents Neutral. The differences between the first two districts and the last two districts are shown in Table 2, i.e., the varying nature of their conditional attributes (d), (e) and (f).

Attributes and classification of conditional attributes are selected via a deductive method, i.e., a long session of interviews was conducted with PRCS experts. PRCS officials suggested a number of attributes, but the given nine were determined to be of top priority. Data is quantified by assigning 1 to High, 2 to Medium, 3 to low and 4 to neutral characteristics. Two districts (Hangu and Lower Dir) have nearly identical classifications, as do Nowshehra and Charsadda, depending on the nature of the area. A maximum of 80 sites were considered in all four districts, and all sites were mapped with conditional attributes, their classifications and decision levels as shown in Tables 1 and 2. Distribution of food took place in these 80 sites, provided by the top management of PRCS. According to Table 3, the decision levels are selected as the highest priority, from high to low. A total of 9 attributes form a to i; the corresponding decision levels are taken as mod () of the 9 given attributes.

5. Algorithms used to generate rules and evaluation measures

Data mining techniques use many algorithms to generate decision rules from the given data. In this study, four different algorithms are applied to generate rules: the exhaustive algorithm [7], genetic algorithm [9, 25], covering algorithm [11] and LEM 2 algorithm [9, 11]. The following evaluation measures were then used to evaluate the results.
  Sensitivity = TP/TP+FN, also called "recall"
  Specificity = TN/TN+FP, where 1-specificity = FP/TN+FPPrecision = TP/(TP + FP)

Table 1
Conditional attributes of Hangu and Lower Dir

Conditional Attributes	Classification	Decision
(a) Access to community	1- District authorities	H
	2- Local community	M
	3- Local volunteers	L
	4- Beneficiary committee	N
(b) Access to site	1- Through main roads	H
	2- Easy to access sites by minor roads	M
	3- Accessible for beneficiaries and PRCS relief workers	L
	4- Transportation	N
(c) Acceptance/ legal base of organization	1- Acceptance from district authorities	H
	2- Armed opposition groups	M
	3- Local/host community	L
	4- Beneficiary committee	N
(d) Security of the area	1- Militancy	H
	2- Ongoing military operations	M
	3- Lack of awareness in beneficiaries	L
	4- Sectarian violence	N
(e) Hazards/risks	1- Armed opposition groups	H
	2- Lack of cooperation among beneficiary and local community	M
	3- Weather condition	L
(f) Seasonal impacts	1- Working hours	H
	2- Rainy season	M
	3- Religious ceremonies	L
(g) Economic considerations	1- Location of site close to main road	H
	2- Beneficiaries' involvement in distribution	M
	3- Transportation	L
	4- Involvement of local volunteers	N
(h) Distribution cycle	1- Cycle scheduling	H
	2- Logistical arrangements	M
	3- Information sharing with beneficiaries and authorities	L
(i) Geographical location	1- Acquisition of appropriate location preferably free of cost	H
	2- Settlement of legal issues/requirements	M
	3- Spacious and plain area	L

Table 2
Conditional attributes of Nowshehra and Charsadda

Conditional Attributes	Classification	Decision
(d) Security of the area	1- Lack of awareness in beneficiaries	H
	2- Local community	M
	3- Ongoing military operations	L
	4- Militancy	N
(e) Hazards/risks	1- Flood	H
	2- Lack of cooperation among beneficiary and local community	M
	3- Weather condition	L
(f) Seasonal impacts	1- Flood	H
	2- Rainy season	M
	3- Working hours	L

Accuracy = (TP + TN) / (TP + FN) + (TN + FP)

Misclassification Error (MisErr) = 1 – Accuracy

Type-I Error = 1–specificity = FP/(FP+TN)

Type-II Error = 1–sensitivity = FN/(TP+FN)

F-measure = 2.(Precision.Recall)/(Precision+Recall)

COV = (number of cases satisfying Condition and Decision)/(number of cases satisfying Decision)

6. Results and discussion of rough set theory

6.1. Results algorithms applied

Four algorithms are applied to the given dataset which include the exhaustive, genetic, covering and LEM2 algorithms. Results are calculated and thoroughly analyzed.

Table 3
Comparison of algorithms

ALGOs	TP	FP	FN	TN	COV	PRE	REC	ER	ACC
Genetic	15.8	0.2	0.2	31.8	1	0.987	0.987	0.012	98%
Exhaustive	15.6	0.4	0.4	31.6	1	0.975	0.975	0.025	97%
Covering	15.4	0.6	0.6	31.4	1	0.962	0.962	0.037	96%
LEM2	14.6	1.4	1.4	30.6	1	0.912	0.912	0.087	91%

Comparison of all algorithms is shown in Table 3. Based on resultant analysis, the genetic algorithm is selected for the proposed dataset because it demonstrates the best accuracy in predicting the result by considering 100% data.

6.1.1. Reducts and rules

Two reduct sets are calculated, and a total of 40 sites are selected by applying rough set theory about the given data. A total of 482 rules are obtained, among which 222 rules support decision 1, (High), 147 rules support decision 2 (Medium) and 113 rules support decision 3 (Low). The rules take the form of "IF A then B". It is not possible to show all rules related to this paper due to space limitations, but rules are selected in descending order as shown in Table 4.

Although rough set theory provided the best overall result, which indicates that all sites that satisfy the decision criteria for a designation of "High" should be considered for food distribution, the top-priority site has yet to be determined. This problem is solved by applying the TOPSIS method, as mentioned in section 3.2.

6.2. Results of TOPSIS method

By applying the TOPSIS method steps to the result provided by the rough set approach to 40 sites, it was demonstrated that the TOPSIS method is able to successfully predict the best and worst sites for future food distribution. The results are shown in Table 5. This work takes into account all four districts, and one site is selected to represent the top priority.

Table 4
Generated rules

Rules	Decision	Strength
(Security of the area = High)&(Sessional Impacts = High)&(Geographical location = High)	(Decision = High)	22
(Security of the area = High)&(Hazards/Risks=High)&(Geographical location = High)	(Decision = High)	18
(Access to community = High)&(Security of the area = High)	(Decision = High)	16
(Access to site = High)&(Sessional Impacts = High)	(Decision = High)	16
(Access to site = Low)&(Acceptance/legal base of organization = Low)& (Security of the area = Low)&(Sessional Impacts = High)	(Decision = Low)	14
(Security of the area = High)&(Sessional Impacts = High)&(attr6 = Medium)	(Decision = High)	14
(Access to site = High)&(Acceptance/legal base of organization = Medium)	(Decision = High)	14
(Access to site = Low)&(Security of the area = High)&(Geographical location = High)	(Decision = High)	14
(Access to community = High)&(Acceptance/legal base of organization = Medium)&(attr6 = Medium)	(Decision = High)	12
(Security of the area = High)&(Distribution = Medium)&(Geographical location = High)	(Decision = High)	12

Table 5
TOPSIS result for 40 sites

Sites	Priority	Sites	Priority	Sites	Priority	Sites	Priority
S78	0.576700728	S68	0.413944532	S34	0.394755358	S30	0.379031331
S80	0.574527344	S18	0.413843895	S2	0.393982109	S9	0.378173552
S79	0.539849739	S25	0.412413341	S14	0.392498378	S40	0.378051893
S77	0.531088902	S63	0.406733933	S12	0.391187388	S8	0.376671271
S76	0.511862922	S65	0.404653161	S10	0.390184843	S15	0.373761641
S57	0.460492984	S62	0.402910406	S21	0.389110907	S20	0.366053234
S35	0.444730396	S31	0.402717182	S61	0.388122395	S50	0.347126307
S71	0.441952049	S59	0.399190461	S33	0.387958645	S48	0.330054896
S26	0.435366112	S70	0.397192116	S7	0.386426001	S58	0.322990182
S38	0.434669949	S11	0.396712345	S3	0.3795342	S49	0.273345739

7. Strengths and weakness

Rough set theory can evaluate the importance of particular attributes and eliminate redundant attributes from a decision table. It can be constructed from minimal subsets of independent attributes, and ensures the same quality of classification as the entire set, i.e., reducts of the set of attributes. Intersections of these reducts provide a core of attributes, which cannot be eliminated without disturbing the ability to approximate the classification, and the generation of logical rules from the reduced decision table. Alternatively, TOPSIS is a simple soft computing method that can be easily implemented using any type of quantitative data.

Along with the above mentioned strengths, there are some weaknesses inherent to these techniques. Selecting appropriate data for these techniques requires extensive study; any addition or deletion can affect the overall accuracy, which may lead to inaccurate results. For site selection, it is expected that the information provided must be accurate and repeatedly revised, leading lead to fewer opportunities for error.

8. Conclusion and future work

This study presents a systematic and flexible approach to deal with rule-based multi-criteria decision making problems, and provides a foundation for decision making regarding appropriate site selection. The successful integration of an RST data mining technique and a multi-criteria decision making method (TOPSIS) has led to successful prediction of sites for PRCS in the four studied districts.

Rough set theory has been used to evaluate the feasibility of different site selection applications, demonstrating 98% accuracy, while the TOPSIS method determines the top priority sites for food distribution. Most PRCS problems involving site selection can be solved according to the proposed method. The advantage of using this method over others is to provide ease, rapid and accurate evaluation of appropriate sites for future food distribution. As rough set theory is a data mining technique, it can demonstrate more accurate results with large changes in data, while the TOPSIS method can be applied to any data.

For future work, variations in rough set theory (i.e., fuzzy rough sets) and in the TOPSIS method (i.e., fuzzy TOPSIS) can also be applied to such data to induce greater accuracy. The evaluation of more sites and attributes are also suggested for incorporation into the suggested technique.

References

[1] B. Walczak and D.L. Massart, Rough sets theory, *Chemometrics and Intelligent Laboratory Systems* **47**(1) (1999), 1–16.

[2] D. Jingfeng and W. Lihai, Research on site selection decision of distribution type logistics park based on the grey correlation TOPSIS method, *Second International Conference on Intelligent System Design and Engineering Application*, Sanya, Hainan, 2012, pp. 890–893.

[3] D.E. Cruz and W.D. Kabiling, A multi-echelon, multi-product-type, site selection and inventory allocation supply chain model for lean facilities, *Journal of the Philippine Institute of Industrial Engineers* **4**(2) (2005), 10–30.

[4] D.T. Lindgren, Site selection issues, in: Land Use Planning and Remote Sensing, Springer Netherlands, 1985, pp. 143–156.

[5] G.K. Chen, Y. Wang and Y. Zhang, Method and model on site selection of aviation spare distribution centre of military aircraft, *IEEE 18th International Conference on Industrial Engineering and Engineering Management (IE&EM)*, Changchun, China, 2011, pp. 1109–1112.

[6] H.H. Lo, Site selection criteria for coastal tourism facilities, *International Symposium on Computer, Consumer and Control*, Taichung, Taiwan, 2012, pp. 890–893.

[7] H.S. Nguyen and S.H. Nguyen, Analysis of STULONG data by rough set exploration system (RSES), *Proceedings of the ECML/PKDD Workshop*, P. Berka, Czech Republic, 2003, pp. 1–82.

[8] J. Komorowaski, Z. Pawlak, L. Polkowski and A. Skowron, Rough sets: A tutorial, *in: Rough Fuzzy Hybridization: A New Trend in Decision-Making*, S.K. Pal and A. skowron, eds., Springer Verlag, Singapor, 1999, pp. 3–98.

[9] J.G. Bazan, H.S. Nguyen, S.H. Nguyen, P. Synak and W. Jakub, Rough set algorithms in classification problem, *in: Rough Set Methods and Applications*, Physica-Verlag HD, Heidelberg, Germany, 2000, pp. 49–88.

[10] J.T. Yao and J.P. Herbert, Financial time-series analysis with rough sets, *Applied Soft Computing* **9**(3) (2009), 1000–1007.

[11] J.W. Grzymala-Busse, A new version of the rule induction system LERS, *Fundamenta Informaticae* **31**(1) (1997), 27–39.

[12] K. Kaneiwa, A rough set approach to multiple dataset analysis, *Applied Soft Computing* **11**(2) (2011), 2538–2547.

[13] K. Yoon and H.K. Ching-Lai-Yoon, *System selection by multiple attribute decision making*, Ph. D. Dissertation, Kansas State University, Manhattan, Kansas, 1980.

[14] K. Yoon and H.K.C. Ching-Lai-Yoon, TOPSIS (technique for order preference by similarity to ideal solution)- A multiple attribute decision making, to be published, 1980.

[15] M. Arabani and M.A. Lashteh Nashaei, Application of rough set theory as a new approach to simplify dams' location, *Scientia Iranica* **13**(2) (2006), 152–158.

[16] M. Vafaeipour, S.H. Zolfani, M.H.M. Varzandeh, A. Derakhti and M.K. Eshkalag, Assessment of regions priority for implementation of solar plants in Iran: New application of a hybrid multi-criteria decision making approach, *Energy Conversion and Management* **86** (2014), 653–663.

[17] M. Yasser, K. Jahangir and A. Mohammad, Earth dam site selection using the analytic hierarchy process (AHP): A case

study in the west of Iran, *Arabian Journal of Geosciences* **6** (2013), 3417–3426.

[18] M.A. Lashteh Neshaei and M. Pirouz, Rough sets theory in site selection decision making for water reservoirs, *Computational Methods in Civil Engineering* **1**(1) (2010), 85–94.

[19] P. Dey, S. Dey, S. Datta and J. Sil, Dynamic disc reduction using rough sets, *Applied Soft Computing* **11**(5) (2011), 3887–3897.

[20] P. Sun and Y. Qian, The study on site selection of passenger station in the valley-city considering negative effects, *Second International Conference on Intelligent Computation Technology and Automation*, Changsha, Hunan, 2009, pp. 841–844.

[21] R. Elhadary and A. Elashry, New approach for landfill site selection using spatial rough set, *Computer Engineering Conference (ICENCO), Seventh International*, Giza, Egypt, 2011, pp. 0–55.

[22] S. Rissino and G. Lambert-Torres, Rough set theory–fundamental concepts, principals, data extraction, and applications, *in: Data Mining and Knowledge Discovery in Real Life Applications*, I-Tech Education and Publishing, 2009, p. 438.

[23] S.H. Zolfani, M.H. Aghdaie, A. Derakhti, E.K. Zavadskas and M.H.M. Varzandeh, Decision making on business issues with foresight perspective; an application of new hybrid MCDM model in shopping mall locating, *Expert Systems with Applications* **40**(17) (2013), 7111–7121.

[24] T.B. Lawson, Recirculating agriculture systems, *in: Fundamentals of Aquacultural Engineering*, Springer, New York, 1995, pp. 192–247.

[25] W. Jakub, Genetic algorithms in decomposition and classification problems, *in: Rough Sets in Knowledge Discovery 2*, Physica-Verlag HD, Warsaw, Poland, **19** (1998), 471–487.

[26] W. Zhang, W. Wu, J. Liang and D.Y. Li, *Theory and Methods of Rough Sets*, Science and Technology Press, Beijing, 2001.

[27] W.J. Jing, Research on site selection of logistics park based on fuzzy comprehensive evaluation method, *Second International Conference on Computer Engineering and Applications*, Bali Island, 2010, pp. 44–47.

[28] Y. Yiyu and Z. Yan, Discernibility matrix simplification for constructing attribute reduces, *Information Sciences* **179**(5) (2009), 867–882.

[29] Z. Pawlak, Rough set theory and its application, *Journal of Telecommunication and Technology* **3** (2002), 7–10.

[30] Z. Pawlak, Rough sets, *International Journal of Computer and Information Sciences* **11**(5) (1982), 341–356.