Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 31 No. 2s (2024)

Rough Set Analysis for Categorizing Motorcycles based on 100 Cubic Centimeters (CC) Engine Displacements

[1]R. Venmani, [2,*]K. Rajendran

[1]Research Scholar, Vels Institute of Science Technology and Advanced Studies, Chennai, Tamil Nadu, India.

[2]Associate Professor, Vels Institute of Science Technology and Advanced Studies, Chennai, Tamil Nadu, India

[1] Email id: rvenmani@yahoo.co.in

[2] Corresponding author: gkrajendra59@gmail.com

Article History:

Received: *01-03-2024*

Revised: *19-04-2024*

Accepted: *15-05-2024*

Abstract:
Rough Set Theory (RST) is a mathematical approach used for dealing with uncertainty and vagueness in decision-making and data analysis. It provides a framework for classifying objects into different equivalence classes based on their attributes or characteristics. In RST, the concept of different cc bikes can be analyzed based on their attributes or characteristics. Each bike can be represented as an object with a set of attributes such as engine displacement, weight, top speed, fuel efficiency and price. Another class may consist of bikes with lower engine displacement, lighter weight, and better fuel efficiency, which could be more suitable for daily commuting. By applying RST, we can analyze the relationship between these attributes and determine the essential and non-essential features of 100 CC bikes. This analysis can help in decision making processes, such as choosing the right bike based on specific requirements or preferences. It's important to note that the application of RST to100 cc bikes is just one example of how this mathematical approach can be used in decision-making and data analysis. The specific attributes and classes may vary depending on the context and purpure of the analysis.

.**Keywords**: Rough set, motorbike, energy, decision-making, reduct

1. Introduction

In decision-making and data analysis, rough set theory provides a mathematical framework for handling ambiguity and uncertainty. Pawlak first presented it as a method of handling imprecise and imperfect data in the early 1980s [1]. The concept of approximations is the foundation of rough set theory. Separating the data into lower and upper approximations offers a formal technique for its analysis and classification. The lower approximation represents the set of things that unquestionably belong to one idea, while the higher approximation represents the set of potential concepts.

The lower and higher approximation operators in Pawlak rough set theory [2] are based on equivalence relations. The need for an equivalence relation in Pawlak rough set models, however, appears to be a highly rigorous requirement that can constrain the applicability of the rough set models. Based on the comparable classes produced by the attribute values, rough set theory divides into three regions: boundary, lower approximation, and upper approximation. According to the data gathered, the upper approximation comprises all the things that can be classified presumably, while the lower approximation contains all the objects that are classified with certainty. The boundary is the difference between the upper and lower approximations [3].

From the perspective of decision-making, an attained rule results in an optimistic decision. In multi-granulation rough set theory, this model is known as the optimistic rough set model. Additionally, Qian [4,5] defines another model known as the pessimistic multi-granulation rough set model. Equivalence relations are expanded to comparable relations or generic binary relations in the extension of the classical rough set model to the generalized rough set model [6]. Data distribution-based rough set models, including the decision-theoretic, the probabilistic, the cloud rough set model, etc [7-9].

It is essential to have a technique for making decisions that will help in identifying and choosing the best motorized vehicles. There are a number of approaches for creating a decision support system (DSS), and one method that may be utilized for decision making systems is WP (Weighted Product) [10]. Although bikes with small engines spend less fuel (per person-kilometer) than cars, the situation may be the opposite for motorcycles with larger engines that are more potent. Additionally, if short non-motorized journeys are replaced with longer motorbike excursions, the overall energy efficiency of the transportation system declines. Pfaffenbichler and Circella [11] analyzed the conditions under which motorbikes can significantly contribute to the development of an energy-efficient and sustainable transportation system.

Butalia et al. [12] revealed that two innovative methods are implemented in Java 1.5 to determine, in light of the relative attribute dependency, the best reductions of condition attributes. The first algorithm provides a simple reduct, while the second one provides a reduct with minimum attributes. In order to create the core of the attribute set or the effective reduct set, unnecessary attributes are removed using Vashist's [13] suggested algorithm. The space complexity and computed time of the suggested approach is compared to that of other known algorithms. It is proven that the suggested method reduces computation time and space without sacrificing the efficacy and quality of the output. Wu and Mi [14] investigated rough sets' mathematical structure in infinite universes of discourse.

Nowadays, there are many various types of motorcycles, including mopeds with a 50 cubic centimeter engine, scooters with engines between 50 and 250 cubic centimeters, and motorbikes with engines up to 1,000 centimeters and even more. Based on these characteristics, we may divide bikes into various equivalence classes in this study using RST. Also, we discover the important and optional aspects of various CC motorbikes by using RST and examining the relationship between these attributes. The results of this research can be useful for making decisions, such as selecting the best bike for a particular set of requirements or desires.

Let's consider that we have a list of bikes with the following attributes:

1. Brand (B): The brand of the bike (e.g., Honda, Yamaha, Suzuki, etc.) as $(x_1, x_2, x_3, x_4 \ldots)$
2. Engine Displacement (E): The engine displacement in cubic centimeters (e.g., 100 cc, 150 cc, etc.).
3. Fuel Efficiency (F): The fuel efficiency of the bike in kilometers per litre (e.g., 40 km/l, 50 km/l, etc.).
4. Price (P): The price of the bike in a certain currency (e.g., USD, INR, etc.).
5. Style (S): The style of the bike (e.g., commuter, sport, cruiser, etc.).

2. Rough Set Analysis

Now, let's apply a simplified rough set analysis with seven types of analysis:

2.1 Lower Approximation (Certain Knowledge)

Identify the bikes that definitely belong to the 100 cc category based on their engine displacement of a set X with respect to S. The lower approximation is represented by $S_*(x)$.

$$S_*(X) = \{x: S(x) \subseteq X\}.$$

2.2 Upper Approximation (Potential Knowledge)

Identify the bikes that potentially belong to the 100 cc category based on their engine displacement X with respect to S. The upper approximation is referred by $S^*(x)$.

$$S^*(X) = \{x: S(x) \subseteq X\}.$$

2.3 Negative Region (Contradiction)

Identify attributes of the set of objects that are contradictory to the 100 cc category of the set X.

$$U - S^*(X)$$

2.4 Boundary Region (Uncertainty)

Identify attributes that are ambiguous and contribute to the uncertainty of bike categorisation of a set X with respect to S and is denoted by $SN_S(X)$.

$$SN_S(X) = S^*(X) - S_*(X).$$

2.5 Indiscernibility Relation (Equivalence)

Identify bikes that are indiscernible (indistinguishable) based on the selected attributes.

Let $R = (A, B)$ be an information system, and $X \subseteq A$. A binary relation $IND_R(X)$ defined in the following way

$$IND_R\ (X) = \{(x_1, x_2) \in A^2 \mid \forall aX, a(x_1) = a(x_2)\}$$

is called a X-indiscernibility relation $[x]_R$. If $(x_1, x_2) \in IND_R(X)$, then x_1 and x_2 are indiscernible (or indistinguishable) by attributes from X.

2.6 Reduct (Minimal Set of Attributes)

Identify the smallest set of attributes that are sufficient to determine the category of a bike.

Using these analyses, you would analyze the attributes of the bikes in your list and determine their relationships to the 100 cc category in Table 1. This analysis would provide insights into the characteristics of bikes that belong to the 100 cc category and help in classifying them based on these attributes.

Please note that this example is a simplified illustration and may not reflect the full complexity of applying rough set theory to real-world data. In practice, rough set analysis involves more rigorous mathematical concepts and techniques.

Table 1. Analysis of engine, power, torque and mileage in 100cc bikes

BIKE (100 CC)	**ENGINE (cc)**	**POWER**	**TORQUE**	**MILEAGE**	**Decision Attributes**
x_1	97.2	8.02	8.05	70	Good
x_2	97	8.02	8.05	70	Good
x_3	97.2	8.02	8.05	70	Good
x_4	97.2	8.02	8.05	70	Good
x_5	102	7.9	8.3	70	Good
x_6	109.7	8.29	8.7	70	Good
x_7	109.51	8.79	9.3	65	Good
x_8	109.51	8.79	9.3	60	Bad
x_9	109.7	8.19	8.7	83.09	Good
x_{10}	109.7	8.19	8.7	73.68	Good
x_{11}	113.2	9.15	9.79	0	Bad
x_{12}	113.2	9.15	9.89	0	Bad

$A = \{x_1, x_2, x_3, x_4, \dots, x_{12}\}$

$B = \{\text{ENGINE, POWER, TORQUE, MILEAGE}\}$

$X_{\text{ENGINE}} = \{97.2, 97, 102, 109.7, 109.51, 113.2\}$

$X_{\text{POWER}} = \{8.02, 7.9, 8.29, 8.79, 8.19, 9.15\}$

$X_{\text{TORQUE}} = \{8.05, 8.3, 8.7, 9.3, 9.79, 9.89\}$

$X_{\text{MILEAGE}} = \{60, 65, 70, 73.68, 83.09\}$

$X_{\text{DECISION}} = \{\text{Good, Bad}\}$

Consider $P = \{\text{ENGINE}\}$

$P = \text{elementary sets}$

$[x_1]_P = [x_3]_P = [x_4]_P = \{x_1, x_3, x_4\}$

$[x_2]_P = \{x_2\}$

$[x_5]_P = \{x_5\}$

$[x_6]_P = [x_9]_P = [x_{10}]_P = \{x_6, x_9, x_{10}\}$

$[x_{11}]_P = [x_{12}]_P = \{x_{11}, x_{12}\}$

The Equivalence classes are

$X_1 = \{x_1, x_3, x_4\}$

$X_2 = \{x_2\}$

$X_3 = \{x_5\}$

$X_4 = \{x_6, x_9, x_{10}\}$

$X_5 = \{x_{11}, x_{12}\}$

$\text{IND(ENGINE)} = \{X_1, X_2, X_3, X_4, X_5\} = \{\{x_1, x_3, x_4\}, \{x_2\}, \{x_5\}, \{x_6, x_9, x_{10}\}, \{x_{11}, x_{12}\}\}$

$$S_*(X) = \{x_2\}$$

$$S^*(X) = \{x_7, x_8\}$$

$$SN_S(X) = S^*(X) - S_*(X) = \{x_7, x_8\}$$

$$U - S^*(X) = \{x_1, x_3, x_4, x_5, x_6, x_9, x_{10}, x_{11}, x_{12}\}$$

This scatter plot visually represents the bikes' data points in the "ENGINE (cc)" vs. "POWER" space. The placement of the data points can give you an idea of how the bikes are distributed in this attribute space. This concept can be extended to additional attributes or even use advanced visualization techniques to represent decision boundaries, discernibility, core, and reducts. However, visualizing rough set theory concepts comprehensively requires advanced data visualization techniques and tools.

Rough set theory involves mathematical concepts and notations that can be challenging to represent in a single diagram. However, we can provide you with a simplified algorithm 1 that illustrates the basic idea of rough set theory and its components: universe of discourse, attributes, and approximations.

Algorithm 1 (Python)

Step1: Find the decision attribute.

Step 2: Create the Concept class based on the decision attribute, identify all the tuples associated with that attribute.

Step 3: Select the Condition attributes. Determine the lower approximation i.e. indiscernible with respect to the selected attributes.

Step 4: Estimate the upper approximation i.e. the objects in the boundary region are indiscernible.

Step 5: Find the rough set

Step 6: Identify the reduct (minimal subset of relevant attributes)

3. Implementation of Proposed Algorithm

Consider implementing the aforementioned technique using a table-based information system. The condition and decision attributes make up the information system's column. The values that the condition and decision attribute can take from the discourse universe are represented by the rows of the information system. Table 2 shows that the universe consists of twelve elements in total. The condition attributes' value sets consist of:

$a = X_1 = \{x_1, x_3, x_4\}$

$b = X_2 = \{x_2\}$

$c = X_3 = \{x_5\}$

$d = X_4 = \{x_6, x_9, x_{10}\}$

Decision Attribute can take values:

$e_0 = X_5 = \{x_8, x_{11}, x_{12}\}$

Table 2. Information System

U	a	b	c	d	e
x_1	1	0	0	0	1
x_2	0	1	0	0	1
x_3	1	0	0	0	1
x_4	1	0	0	0	1
x_5	0	0	1	0	1
x_6	0	0	0	1	1
x_7	0	0	0	0	1
x_8	0	0	0	0	0
x_9	0	0	0	1	1
x_{10}	0	0	0	1	1
x_{11}	0	0	0	0	0
x_{12}	0	0	0	0	0

Using our approach, the first condition attribute, "a," is removed from table 2 and the consistency of the remaining table, table 3, is checked. When we have identical options for two or more rows or cases with the same values of condition characteristics, a table is considered to be consistent; if not, it is inconsistent.

Table 3. Attribute ' a ' is eliminated

U	b	c	d	e
x_1	0	0	0	1
x_2	1	0	0	1
x_3	0	0	0	1
x_4	0	0	0	1
x_5	0	1	0	1
x_6	0	0	1	1

x_7	0	0	0	1
x_8	0	0	0	0
x_9	0	0	1	1
x_{10}	0	0	1	1
x_{11}	0	0	0	0
x_{12}	0	0	0	0

Table 3 is created by removing the condition attribute ' a ' from Table 2.

Upon examining table 3, it becomes evident that the removal of attribute ' a ' from table 2 results in inconsistency since, for each of the following values of the condition attributes: ' b', ' c ', ' d ' the value of the decision attribute, ' e ', differs for x_1, x_3, x_4 and x_8, x_{11}, x_{12}.

$$x_1, x_3, x_4: b_0 c_0 d_0 \rightarrow e_1$$
$$x_8, x_{11}, x_{12}: b_0 c_0 d_0 \rightarrow e_0$$

Thus, CORE is used with condition attribute' a '.
Table 4 remains after removing characteristic ' b ' from Table 2.

Table 4. Attribute ' b ' is eliminated

U	a	c	d	e
x_1	1	0	0	1
x_2	0	0	0	1
x_3	1	0	0	1
x_4	1	0	0	1
x_5	0	1	0	1
x_6	0	0	1	1
x_7	0	0	0	1
x_8	0	0	0	0
x_9	0	0	1	1
x_{10}	0	0	1	1
x_{11}	0	0	0	0
x_{12}	0	0	0	0

Upon examination, Table 4 appears to be devoid of any discrepancies. Eliminating the characteristic ' b ' results in no inconsistency. Hence, according on the information, attribute ' b ' is not a CORE.

Reduct $1 = \{\mathrm{a}, \mathrm{c}, \mathrm{d}\}$

Table 5 remains after removing attribute ' c ' from Table 2.

Table 5. Attribute ' c ' is eliminated

U	a	b	d	e
x_1	1	0	0	1
x_2	0	1	0	1
x_3	1	0	0	1
x_4	1	0	0	1
x_5	0	0	0	1
x_6	0	0	1	1
x_7	0	0	0	1
x_8	0	0	0	0
x_9	0	0	1	1
x_{10}	0	0	1	1
x_{11}	0	0	0	0
x_{12}	0	0	0	0

The removal of the attribute ' c ' doesn't result in inconsistent data. As a result, condition characteristic ' c ' is not utilized as the core. However, Table 5's remaining condition attribute is regarded as reduct.

Hence,

$\text{Reduct } 2 = \{\text{a}, \text{b}, \text{d}\}$

Table 6. Attribute ' d ' is eliminated

U	a	b	c	e
x_1	1	0	0	1
x_2	0	1	0	1
x_3	1	0	0	1
x_4	1	0	0	1
x_5	0	0	1	1
x_6	0	0	0	1
x_7	0	0	0	1
x_8	0	0	0	0
x_9	0	0	0	1
x_{10}	0	0	0	1
x_{11}	0	0	0	0
x_{12}	0	0	0	0

There is no inconsistency in the dataset when condition attribute ' d ' is removed from table 6.

As a result of this, attribute ' d ' is not a CORE.

$\text{Reduct } 3 = \{\text{a}, \text{b}, \text{c}\}$

Characteristic ' a ' is the only essential characteristic. ' a ' also appears in each reduct set.

Additionally, we understand that

Core $= \cap$ Reducts

$\mathrm{CORE}\ (\mathrm{C}) = \cap$ {{ Reduct 1}{ Reduct 2}{ Reduct 3}}

and the Reduct sets are

Reduct $1 = \{\mathrm{a}, \mathrm{c}, \mathrm{d}\}$

Reduct $2 = \{\mathrm{a}, \mathrm{b}, \mathrm{d}\}$

Reduct $3 = \{\mathrm{a}, \mathrm{b}, \mathrm{c}\}$

$\mathrm{CORE}\ (\mathrm{C}) = \{\mathrm{a}\}$

This further demonstrates that our algorithm is accurate.

4. Conclusion

The application of rough set theory to the task of finding the best bike in different cubic centimeters (cc) has provided valuable insights in this paper. By employing rough sets, we can effectively analyze and categorize bikes based on various attributes and features relevant to different cc categories. This approach helps us to identify the essential attributes that contribute to the superiority of a bike within a specific cc range. Through the reduction of redundant information and the extraction of essential characteristics, rough set theory streamlines the decision-making process for selecting the optimal bike in each cc category. Overall, utilizing rough set theory enhances our ability to make informed choices while considering the intricate relationships between attributes and bike performance in different cubic centimeter ranges.

References

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

[2] Pawlak, Z. Rough Sets-Theoretical Aspects of Reasoning About Data, *Kluwer Academic Publishers*, Boston, MA, 1991.

[3] Thangavel, K., & Pethalakshmi, A. (2009). Dimensionality reduction based on rough set theory: A review. *Applied soft computing*, *9*(1), 1-12.

[4] Qian, Y., Liang, J., Yao, Y., & Dang, C. (2010). MGRS: A multi-granulation rough set. *Information sciences*, *180*(6), 949-970.

[5] Qian, J., Han, X., Yu, Y., & Liu, C. (2023). Multi-granularity decision-theoretic rough sets based on the fuzzy T-equivalence relation with new strategies. *Journal of Intelligent & Fuzzy Systems*, (Preprint), 1-15.

[6] Liang, D., Liu, D., Pedrycz, W., & Hu, P. (2013). Triangular fuzzy decision-theoretic rough sets. *International Journal of Approximate Reasoning*, *54*(8), 1087-1106.

[7] Yao, Y., & Yao, B. (2012). Covering based rough set approximations. *Information Sciences*, *200*, 91-107.

[8] Yao, Y. (2011). The superiority of three-way decisions in probabilistic rough set models. *Information sciences*, *181*(6), 1080-1096.

[9] Raghavan, R., & Tripathy, B. K. (2011). On some topological properties of multigranular rough sets. *Advances in Applied Science Research*, *2*(3), 536-543.

[10] Jupon, R. M., Wati, R., Kristina, M., & Nagara, E. S. (2023). Determining the Purchase of Used Motorcycles Using the WP Method Based on Mobile Web. *International Journal Of Multidisciplinary Research*, *1*(1), 1-12.

[11] Pfaffenbichler, P., & Circella, G. (2009). The role of motorized 2-wheelers in an energy efficient transport system. *Proc. of the European Council for an energy efficient economy, Le Colle Sur Loup, Côte d'Azur*, 1345-1354.

[12] Butalia, A., Dhore, M., & Tewani, G. (2008). Applications of rough sets in the field of data mining. In *2008 First International Conference on Emerging Trends in Engineering and Technology* (pp. 498-503). IEEE.

[13] Vashist, R. (2015). An algorithm for finding the reduct and core of the consistent dataset. In *2015 International Conference on Computational Intelligence and Communication Networks (CICN)* (pp. 741-745). IEEE.

[14] Wu, W. Z., & Mi, J. S. (2011). Some mathematical structures of generalized rough sets in infinite universes of discourse. In *Transactions on Rough Sets XIII* (pp. 175-206). Springer Berlin Heidelberg.