Criteria Selection Decision Making of Hotels through Rough Set Theory

Haresh Kumar Sharma\textsuperscript{a}, Samarjit Kar\textsuperscript{b}

Department of Mathematics, National Institute of Technology Durgapur, West Bengal, 713 209, India
\textsuperscript{a,b}E-mail address: hareshshrm@gmail.com, dr.samarjitkar@gmail.com

ABSTRACT

This study proposes Rough Set Theory (RST) based Decision Making for evaluating the performance of the Delhi hotels. In this study, we use rough set theory to provide a set of the decision rule and important feature sets, which can help hotel management to improve the quality of the hotel to attract maximum tourists. This study shows that RST helps in identifying tourist, identifying their specialty and facilitating the development of hotels strategies. This research can helps hospitality management to understand traveller requirement and enhance the service best of the hotel industry.

\textbf{Keywords:} Decision making, Hotels, Rough Set Theory, Tourist

1. INTRODUCTION

In India, the tourism sector is an essential source of earnings, foreign exchange, and employment. It contributes to 6.23\% to the National Gross Domestic Product (GDP), annual growth rate of 14.12\% of foreign exchange to the Indian economy and about 9 \% jobs in the country. According to tourism completion report (2013), Indian is at 11th position in Asia and 65th in world travel index. Hotel management is necessary for policy, decision makers, and customer satisfaction. Analyzing the quality of the hotel is extremely helpful for making travel decisions and policymaking. Quality and Relation study can provide dependable and valuable data for the hospitality industry for future planning of journey.
Statistical analysis and data mining approaches have a good impact on the planning of hotel’s activities. Using Statistical analysis and data mining approach defines the correlation between criteria. Therefore, to focus on the issue of Hotel Quality Analysis, an empirical study of the Indian hotel industry has been selected as the subject of this research [1-5].

Over the past many years’ different techniques have been applied in hotel’s data analysis. Many studies have analyzed the quality of the hotels using descriptive statistics, multiple regression techniques, factor analysis, principle component technique. In [23] analyzed perceptions of students of the quality aspects of Virtual Field Trip (VFT) technology-based education and teaching equipment.

In [7], examined Customer Experience with Budget Hotels using exploratory factor analysis. Xu and Li [19] applied a text mining approach in customer’s satisfaction or dissatisfaction data of hotels. In [13] applied the fuzzy Delphi approach and fuzzy analytic hierarchy approach to analyse and evaluate the criteria of Taiwan's hotel.

In [6] analyzed the service quality of hotel data using regression analysis. A new approach rough set theory is helpful for discovering the relationships among dependent and independent variables. Rough set theory can be easily applied to qualitative data. In uncertain and vague dataset RST has been employed for creation of decision rules [13]. In [14] applied RST to review customer’s evaluations of a destination. Moreover, the RST can really be implemented to units labelled by immaterial facts in which statistical tools fail to offer fruitful consequences [8-12].

In this research, we have applied rough set theory to exploring the quality analysis of hotel industry. The primary goal of the have a look at is to research the hotel best the usage of specific variables using rough set method. In this research, we have obtained the necessary criteria using attribute reduction technique applied on hotel data. The rest of the paper is structured as follows.

Section 2 introduces the basic concepts of rough set theory. The next section describes the data for the analysis. The following presents the empirical findings. Next section describes the discussion and the policy implications of hotel management and the final section summarize the conclusions [15-23].

2. ROUGH SET THEORY

The rough set is a new mathematical technique to handle imprecision, vagueness, and uncertainty (Pawlak, 1982). For analysis of imprecise description of objects, RST is an excellent mathematical method. The adjective vague describe the information system that is ambiguity or uncertainty that chase from statistics granulation. The same information in data set form similar relation.

The indiscernibility (similar) relation evolved is that this way is a mathematical basis of the rough set theory. RST induces a separation of the objects of the universal set into part of indiscernible (similar) members, called elementary set. The primary idea of the rough set concept is based totally on indiscernible (IND) relation. Using indiscernibility relation RST gives two approximations set, called lower and upper approximation of an information set.

Let U be a universal set of objects contains Y be the non-empty finite set of objects. A be a nonempty finite set of criteria, then P = (Y, A) is called a dataset where A contains two subsets C, D, where C and D are condition and decision attribute, respectively.
For \( P = (Y, A) \) and \( S \subseteq A, R \subseteq Y \) can be approximated based on the knowledge contains two approximation set \( S \)-lower and \( S \)-upper approximation of \( R \), denoted by \( S(R) \) and \( \tilde{S}(R) \) respectively; where

\[
S(R) = \{ y | [y]_S \subseteq R \}
\]

\[
\tilde{S}(R) = \{ y | [y]_S \cap R \neq \emptyset \}
\]

The members in \( S(R) \) is known as the set of all members of \( Y \) that can be certainly described as a member of \( R \) in the knowledge \( S \) whereas members in \( \tilde{S}(R) \) is the set of all elements of \( X \) that can be possible described as a member of \( R \). The boundary region of set \( Y \) is denoted as: \( BN_S(R) = \tilde{S}(R) - S(R) \) is the set of objects which cannot strictly categorize into \( R \) contains knowledge \( P \). If the boundary region does not contain any element, then lower approximation (LA) and upper approximation (UA) set are similar. In the reverse case, if the boundary region contains some element than the set \( R \) is known as rough set.

**Attribute Reduction Technique:**

Using attribute reduction is to recognize the necessary criteria and to reduce the unnecessary attribute of the data set. Any information system may contain one or more reduce. The positive region is an important concept of the RST (Pawlak, 1991).

The \( C \)-Positive region of \( D \) having the set of all instance of the universal set that is definitely classified into the separation of \( U/D \) by using knowledge from \( C \). The \( C \)-Positive region of \( D \) defined as:

\[
POS_C(D) = \bigcup_{R \in U/D} CR
\]

If attribute set \( F \subseteq C \) is called a reduct of \( C \) with respect to \( D \) if following conditions are satisfied

\[
(i) POS_C(D) = POS_F(D),
\]

\[
(ii) POS_C(D) \neq POS_{F-\{f\}}(D), for any f \in F.
\]

**The accuracy of approximation and dependency of attributes (Pawlak, 1991):**

The accuracy of approximation (AA) of any subset can be denoted by:

\[
\alpha_S(R) = \frac{|S(R)|}{|\tilde{S}(R)|},
\]

where \( 0 \leq \alpha_S(R) \leq 1 \). If the accuracy of approximation of set is 1 then set is called crisp otherwise set is rough.

The dependency of attributes depends on the object in the lower approximation to all object-in-universe and denoted by:
\[ \gamma_C(D) = \frac{|POS_C(D)|}{|U|} \]

If \( \gamma_C(D) = 1 \) we say that D depends completely on C, and if \( 0 \leq \gamma_C(D) \leq 1 \), we say that D depends partially on C. Furthermore, if \( \gamma_C(D) = 0 \) then D is completely independent from C.

The fundamental concept of RST is the advent of decision rules which is based on if and then statement. Decision rules use to preserve the core interpretation of the attribute set from the provided furnished data of unique hassle that is a further vast factor of RST.

Following steps are used for data table observation.

1. Data collection.
2. Calculation of approximation sets of the universal set.
3. Compute the C-positive region of D.
4. Calculate reduct and core of A.

### 3. CASE STUDY AND RST ANALYSIS

In this study, we considered the case study of Delhi’s hotels and examines the effect of location (LC), hospitality (HT), facilities (FS), cleanliness (CN), value for value (VFM), and food (FD) and price (P) on overall rating (OR). All criteria are used according to the objective of availability and the purpose of the study. In our empirical analysis, all the different criteria of hotels are considered in this analysis.

Data related to hotel industry is collected from the tourism website (https://www.makemytrip.com). Tourism websites are usually reviewed based on customer’s feedback. When new tourist willing to visit any place then they can browse these websites and look at these reviews and make decisions based on the previous review. It is very difficult to make decisions when there is a large number of options available for any of the hotels, tourist sites or services available.

The proposed approach is helpful to the hotel selection suitable for the current scenario of the city. It is generally considered when reviewing the hotel's review, location, service, facilities, price, cleanliness, and food. The proposed method can be used to identify the appropriate hotel on the basis of existing data.

**Table 1. Accuracy and Degree of Dependency**

<table>
<thead>
<tr>
<th>Class</th>
<th>LA</th>
<th>UA</th>
<th>AA</th>
<th>Degree of dependency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>7</td>
<td>15</td>
<td>0.4667</td>
<td></td>
</tr>
<tr>
<td>Very good</td>
<td>20</td>
<td>28</td>
<td>0.7143</td>
<td>0.8330</td>
</tr>
<tr>
<td>Excellent</td>
<td>23</td>
<td>31</td>
<td>0.7419</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Decision Rules from Data Set

<table>
<thead>
<tr>
<th>Rule No.</th>
<th>Decision Rule</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(FS = Excellent) &amp; (CN = Excellent) =&gt; (OR = Excellent)</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>(FS = Excellent) &amp; (VFM = Excellent) &amp; (FD = Fair) =&gt; (OR = Excellent)</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>(HT = Excellent) &amp; (P = Very High) =&gt; (OR = Excellent)</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>(LC = Very Good) &amp; (VFM = Good) &amp; (FD = Fair) =&gt; (OR = Excellent)</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>(LC = Good) &amp; (FD = Poor) &amp; (price = Medium) =&gt; (OR = Very Good)</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>(LC = Good) &amp; (HT = Good) &amp; (FS = Very Good) =&gt; (OR = Very Good)</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>(LC = Good) &amp; (HT = Good) &amp; (FS = Very Good) =&gt; (OR = Very Good)</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>(HT = Very Good) &amp; (price = High) =&gt; (OR = Very Good)</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>(VFM = Good) &amp; (price = High) =&gt; (OR = Very Good)</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>(HT = Very Good) &amp; (CN = Good) =&gt; (OR = Very Good)</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>(FS = Very Good) &amp; (CN = Good) &amp; (price = Low) =&gt; (OR = Very Good)</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>(FD = Very poor) =&gt; (OR = Very Good)</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>(LC = Very Good) &amp; (CN = Very Good) &amp; (FD = Poor) =&gt; (OR = Very Good)</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>(FS = Good) &amp; (VFM = Good) &amp; (price = Low) =&gt; (OR = Good)</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>(FS = Good) &amp; (FD = Fair) &amp; (price = Medium) =&gt; (OR = Good)</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>(LC = Good) &amp; (price = Low) =&gt; (OR = Good)</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>(LC = Excellent) &amp; (VFM = Good) &amp; (price = Medium) =&gt; (OR = Good)</td>
<td></td>
</tr>
</tbody>
</table>

4. RESULT AND DISCUSSION

To obtain decision rules we applied rough set theory in hotel data set. The data set classified into three classes excellent, very good and good. Information system used for RST to obtain the value of degree of dependency and accuracy of approximation. The results of degree of dependency and accuracy of approximation are listed in Table 3. The overall dependency between conditional and decision attributes is 83%. Using attribute reduction technique, we can obtain indispensable set of attributes. Indispensable set is smaller set of attributes which can describes the necessary attribute in decision table. In our analysis, attributes location, facilities,
Cleanliness, food, Price are important attributes. Using Rough set theory in data set, we were analyzed data set and decision rules were obtained. 18 certain decision rules were obtained from the decision set. The ROSE algorithm has used to create decision rules (Predki, et al. 1998).

In 19 decision rules, five rules related to decision excellent, ten rules to decision very good and four rules to decision good. The results of decision rules are presented in Table 4. From Table 4, we can see that, rule 1 identified by two attributes, “facilities” and “Cleanliness”, which means that “If facilities of the hotel are excellent and Cleanliness is fair THEN overall rating of the hotel will be excellent.” The support of Rule 1 was 12. According to Rules 2 “If a hotel with excellent facilities, excellent value for money, and quality of food is fair, then the overall rating of the hotel will be excellent.” Rule 2 identified three attributes, “facilities”, “food”, and “value for money” and support of rule 2 was 9.

5. CONCLUSION

The objective of this study was to model rough set for the quality analysis of the hotels. Here seven important condition attribute (e.g., location, hospitality, cleanliness, facilities, value for money, food and value) and one decision attribute overall ranking for the hotel were included for the analysis purpose. The main contribution of the paper is that more precise and consistent decision rules using roughest approach with hotel attributes without any particular statistical assumption. In this analysis, we found that five attributes (location, facilities, Cleanliness, food, Price) are more significant attributes in this data set. This proposed study serves well and produces a satisfactory result.

References


