Int. J. Mach. Learn. & Cyber.
DOI 10.1007/s13042-014-0297-2

ORIGINAL ARTICLE

Linguistic rough sets

Manish Agarwal · Themis Palpanas

Received: 25 December 2013 / Accepted: 4 September 2014

Abstract We introduce linguistic rough set (LRS) by integrating linguistic quantifiers in the rough set framework. The proposed LRS is inspired by the ways in which humans process imprecise information. It operates directly with the linguistic summaries and caters to imprecision implicit in the real world with partial knowledge. The measures of LRS are developed and its properties are investigated in detail. An approach is proposed for approximation of fuzzy concepts with the proposed LRS. This approach is applied in a real world case-study on the credit scoring analysis problem.

Keywords Linguistic rough set · Linguistic quantifier · Fuzzy · Multi criteria decision making · Approximation · Inclusion degree

List of symbols

U	Universe of objects
x_i	Object or alternative
C	Set of attributes/criteria
f	Function
V	Domain of values
IS	Information system
IND	Indiscernibility relation
U/B	Partition of U with respect to attribute set B
$[x_i]$	Equivalence class generated with IND
$\underline{B}(D)$	Lower approximation of set D with respect to attribute set B
$\overline{B}(D)$	Upper approximation of D with respect to attribute set B
$\widetilde{R}(x_i, x_j)$	Fuzzy similarity relation between x_i and x_j
$\mu_{C_j}(x_i)$	Fuzzy membership of x_i in C_j
$\mu_{\underline{\tilde{R}}(Y)}(C_j)$	Lower approximation of membership of fuzzy concept C_j in fuzzy concept Y
$\mu_{\overline{\tilde{R}}(Y)}(C_j)$	Upper approximation of membership of C_j in Y
$\mathcal{F}(U)$	Power set of all fuzzy subsets defined on U
supp$V$	Support of a fuzzy set V
card$V$	Cardinality of a fuzzy set V
S	Linguistic label set
$\mathcal{D}(W, V)$	Degree of inclusion of fuzzy concept V in W
λ	Degree of certainty of approximations
$\underline{C}_\lambda(Y)$	Lower approximation of Y in terms of a set C of fuzzy concepts
$\overline{C}_\lambda(Y)$	Upper approximation of Y in terms of set C of fuzzy concepts
$T_C(Y)$	Quality of approximation of fuzzy concept Y
κ	Crispness coefficient
$\alpha_\lambda(Y)$	Accuracy of approximation of Y with degree of certainty λ

M. Agarwal (✉)
Department of Computer Science and Engineering, Indian Institute of Technology, New Delhi, India
e-mail: magarwal@cse.iitd.ernet.in

T. Palpanas
Department of Information Engineering and Computer Science, University of Trento, Trento, Italy

1 Introduction

Imprecision, incompleteness and vagueness are inherently associated with the real world. Notwithstanding, we conclude and take decisions on the basis of all the information at our disposal and our beliefs and perceptions. This further highlights the need to effectively process the uncertain,

Published online: 10 October 2014

vague and imprecise information. Fuzzy set and rough set [15, 41] theories have been proposed to deal with the imprecision in the information. While, the fuzzy set describes vagueness in terms of membership function (MF), the rough set theory helps to approximate an imprecise concept in terms of crisp available knowledge as lower and upper approximations. The conventional rough set may only be used when we have a crisp information system, and often, it is very difficult to represent the real world imprecision in terms of crisp numbers. Various hybridizations of these two theories of uncertainty have appeared like fuzzy rough sets [16, 18, 32] and rough fuzzy sets [16, 32]. More recently, the conventional rough set theory is extended to the fuzzy environment in [33, 34] in an attempt to improve its performance. Nevertheless, the fact remains that fuzzy set and rough set theories and their hybridizations still operate on numbers (MF).

The perception, subjectivity, attitudes, priorities and incomplete knowledge (of agent) add to the interpretation issues with MF in FRS [29, 30]. While the objective of FRS is to represent imprecision, in many situations it also amounts to quantification of imprecision in terms of membership degrees leading to distortion of original imprecision. In our view the imprecision may be well represented by the same semantics as it is thought or perceived by a human brain. We often resort to granulate the objects, events or situations on basis of one or the other characteristics depending upon the criterion in focus in order to simplify the process of absorbing or analyzing the objects or situations; for example, all small cars, costly cars, efficient cars. Similarly we categorize decision making situations as high risk, high gain; low risk, low gain; high input cost, high risk and high profitability propositions. Granular computing [12, 35] is developed based on granulation of knowledge for complex problem solving. Rough set is extended to multi granulation rough set in [36], in which a target concept is approximated by multiple granulations. It is further studied in [37, 38].

The "computing with words" (CW) methodology has been proposed in [31], in which the objects of computation are words and propositions drawn from a natural language, e.g., small, large, heavy, likely, etc. In a similar vein, the notion of linguistic (fuzzy) quantifier [11, 14] is proposed to indicate imprecision through words, without resorting to numeric membership grades. For example: "most kind men live long". The linguistic quantifier in this representation is "most". Similarly, other quantifiers in the same class could be many, very, several, high, low etc. The natural representations of an imprecise situation are often characterized with such linguistic quantifiers that concern the representation of collections with unclear boundaries by means of a variable whose values are fuzzy sets. A linguistic quantifier can also be regarded as a form of information summarization or granulation.

The granulation of information is inspired by human thought process. The human brain has a remarkable capability to think, assess, summarize, and memorize an imprecise situation in terms of these linguistic quantifiers. Often, such imprecise assessments of key attributes form the foundation of decisions. For example, in a selection interview, the experts vaguely assess a candidate against multiple attributes and then place each attribute of the candidate in a broad bracket (a linguistic quantifier). The final decision of selection is a vague aggregation of these quantifiers, assigned to a candidate against various attributes. In the similar vein, the rough set theory is also based on granulation of information, in which an incompletely known concept is approximated in terms of information granules. This process is very much similar to human's mind way of simplification of knowledge.

Both rough set and CW theories are concerned with granulation of information. Despite these commonalities, the research in these two fields is still disconnected. To some extent, this can be explained by differences in the concrete problem settings considered. Given important commonalities but also differences in terms of methodologies, one can argue that CW can complement the rough set theory and their cross-fertilization could result in interesting uncertainty representation structures. CW is a knowledge representation tool to deal with vagueness implicit in human natural language representations. The rough set theory, on the other hand, is concerned with granulation of available information. What may be specifically interesting about the combination of CW with the rough set theory are the extended approximation capabilities of the combined framework to directly operate with natural representations without any measurements and any computations.

This paper is a concrete realization of exactly this idea that is inspired by the remarkable human cognition and decision process that is able to perform a wide variety of physical and mental tasks without exact measurements, but on the basis of vague assessments and granulation of information. Here, we integrate linguistic quantifiers in the rough set framework to develop linguistic rough set (LRS). LRS approximates an imprecise concept, directly in terms of linguistic quantifiers. The linguistic quantifiers in LRS are akin to imprecise granules of information.

Linguistic rough set, like the conventional rough set, can be also used to discover hidden knowledge in the form of logical decision rules using the available information in terms of the information granules. The proposed LRS is designed to yield robust and generalized approximations that reasonably hold true for a larger population. This potentially makes LRS a prospective choice for application in decision making under uncertainty.

The key contributions of the paper can be summarized as follows:

- Combining CW with the rough set theory to retain the underlying imprecision in the approximations.
- Introduction of linguistic approximation space.
- Investigation of properties of linguistic approximation space.
- Introduction of linguistic rough set.
- Application of linguistic rough set in a decision making application.

The rest of the paper is organized as follows. Section 2 discusses an overview of the related topics. In Sect. 3, we give the motivations for developing linguistic rough set (LRS). Here, we also present the concept of linguistic approximation space. In Sect. 4, we develop linguistic rough set and investigate its measures and properties. Section 5 is concerned with an approach to apply LRS in multi criteria decision making, involving linguistic assessments by experts. The proposed approach is also applied in a real world case-study in the area of oil mining investment. Finally, Sect. 6 gives the conclusions of the study.

2 Preliminaries

The fuzzy set and rough set theories are extensions of classical set theory to deal with vagueness, and imprecision and insufficient knowledge respectively. Though the two theories are different and complimentary to each other, they are yet related in the sense that both the theories address the problem of information granulation. The fuzzy sets are centered upon fuzzy information granulation, whereas the rough set theory is focused on crisp information granulation [17]. A review of the basic concepts of the rough set and fuzzy rough set theories is given here to build the background for the linguistic rough set.

2.1 Rough sets

The rough set method classifies objects of discourse into equivalence classes containing indistinguishable objects with respect to some attributes. The equivalence classes or the knowledge granules form the basic elements to approximate the object sets. The attribute set can be considered as knowledge that helps partition the universe into the basic concepts. Indiscernibility is central to rough set and the methodology to achieve this is quite close to human brain's thought process. Once the indiscernibility in condition concepts (available information) is determined, the resulting equivalence classes are used to approximate the decision concepts (concepts to be approximated). The rough set is studied from the perspective of dependency space and closure system in [40]. The basic concepts of the rough set theory are now discussed.

An approximation space [15] can be defined as a system (U, C), where $U = \{x_1, x_2, \ldots, x_m\}$ and C is the set of attributes (features, criteria). The family of attributes is also known as knowledge in the universe. Each attribute $c \in C$ defines an information function $f_c : U \to V_c$, where V_c is the set of values that attribute c may take, and is called the domain of attribute c. The value that attribute c takes for x_i is denoted as $f_c(x_i)$. An approximation space is also termed as information system (IS).

$$\mathrm{IS} = (U, C). \tag{1}$$

The approximation space forms the framework in which an imprecise concept is approximated in terms of equivalence classes. For every set of attributes $B \subset C$, an indiscernibility relation $\mathrm{IND}(B)$ [15] is an equivalence relation such that two objects, x_i and x_j, are indiscernible by the set of attributes B, if $f_b(x_i) = f_b(x_j)$ for every $b \subset B$.

$$\mathrm{IND}(B) = \left\{ \left(x_i, x_j\right) \in U \times U | \forall b \in B, f_b(x_i) = f_b\left(x_j\right) \right\}. \tag{2}$$

$\mathrm{IND}(B)$ generates a partition of U which is denoted as $U/\mathrm{IND}(B)$ or simply U/B.

$$U/B = \otimes \{ U/\mathrm{IND}(\{b\}) | \forall b \in B \} = \left\{ [x_i]_B : x_i \in U \right\} \tag{3}$$

where, $[x_i]_B$ denotes the equivalence class of $\mathrm{IND}(B)$ and is called elementary set in B because it represents the smallest discernible groups of objects. The approximation of objects using the rough set theory is always in terms of these equivalence classes. $\otimes$ operation is specified as $C \otimes B = \{ X \cap Y | X \in C, Y \in B, X \cap Y \neq \emptyset \}$.

A rough set approximation [15] of an arbitrary non-empty subset D of U is characterized by two unions of elementary sets $\langle \underline{B}(D), \overline{B}(D) \rangle$ referred to as B-lower and B-upper approximations of D over IS (U, B), which are given as

$$\begin{cases} \underline{B}(D) = \left\{ x_i \in U | [x_i]_B \subseteq D \right\} \\ \overline{B}(D) = \left\{ x_i \in U | [x_i]_B \cap D \neq 0 \right\} \end{cases} \tag{4}$$

where, $\underline{B}(D)$ and $\overline{B}(D)$ respectively denote the lower and upper approximations of an imprecise concept Y.

2.2 Fuzzy rough sets

In fuzzy rough set (FRS), an equivalence relation is replaced by a fuzzy similarity relation as a key criterion to granulate the knowledge. Hence, the fuzzy equivalence classes are central to FRS in the same way as crisp equivalence classes are central to rough set [11]. A fuzzy binary relation $\widetilde{R}$ on U is called a fuzzy similarity relation if the properties of reflexivity $\left(\widetilde{R}(x, x) = 1\right)$, symmetry $\left(\widetilde{R}(x, y) = \widetilde{R}(y, x)\right)$, and T-transitivity

$\left(\widetilde{R}(x,z) \geq \widetilde{R}(x,y) \wedge_T \widetilde{R}(y,z)\right)$ hold good. The fuzzy similarity relation leads to a generation of the family of normal fuzzy sets produced by the fuzzy partitioning of the universe of discourse. These fuzzy sets can be viewed as fuzzy equivalence classes, $[x]_{\widetilde{R}}$, and play the same role as the equivalence classes play in the case of rough set.

Similar to an approximation space in the rough set theory, we could conceive fuzzy approximation space [17] over U, corresponding to a similarity relation $\widetilde{R}$ as pair (U, $\widetilde{R}$). A fuzzy equivalence relation $\widetilde{R}$ replaces a crisp equivalence relation in a fuzzy approximation space. In contrast to rough set, where the approximations are in the terms of crisp equivalence classes, the soft fuzzy granules arising from fuzzy equivalence relation are used to approximate a fuzzy concept in FRS. Formally, the concept of FRS can be defined as follows.

Let $\widetilde{R}$ be a fuzzy binary relation on U, C_j be a fuzzy equivalence class, and Y be the concept to be approximated. A fuzzy rough set is a pair of fuzzy sets on U such that for every $x \in U$ [16]

$$\mu_{\underline{\tilde{R}}Y}(C_j) = \inf_i \max\left\{1 - \mu_{C_j}(x_i), \mu_Y(x_i)\right\}, \forall j \tag{5}$$

$$\mu_{\overline{\tilde{R}}Y}(C_j) = \sup_i \min\left\{\mu_{C_j}(x_i), \mu_Y(x_i)\right\}, \forall j \tag{6}$$

where, $\mu_{C_j}(x_i)$ denotes the membership of x_i in C_j, $\mu_Y(x_i)$ denotes the membership of x_i in Y, $\mu_{\underline{\tilde{R}}Y}(C_j)$ denotes the degree of certain membership of C_j in Y, $\mu_{\overline{\tilde{R}}Y}(C_j)$ denotes the degree of possible membership of C_j in Y, and $\left(\mu_{\underline{\tilde{R}}Y}(C_j), \mu_{\overline{\tilde{R}}Y}(C_j)\right)$ is a fuzzy rough set defined with fuzzy equivalence relation $\tilde{R}$.

The definition of fuzzy rough set was modified lately in [18] so that the same degenerates to conventional rough set when all equivalence classes are crisp.

$$\mu_{\underline{\tilde{R}}Y}(C_j) = \sup_{F\in U/B} \min\left(\mu_{C_j}(x_i), \inf_{x_i\in U} \max\left\{1 - \mu_{C_j}(x_i), \mu_Y(x_i)\right\}\right) \tag{7}$$

$$\mu_{\overline{\tilde{R}}Y}(C_j) = \sup_{F\in U/B} \min\left(\mu_{C_j}(x_i), \sup_{x_i\in U} \min\left\{\mu_{C_j}(x_i), \mu_Y(x_i)\right\}\right). \tag{8}$$

If $\mathcal{F}(U)$ denotes the set of all fuzzy subsets from U. Then for $V, W \in \mathcal{F}(U)$ the following relations are true:

$$\mu_{V\cup W}(x_i) = \max\{\mu_V(x_i), \mu_W(x_i)\} \tag{9}$$

$$\mu_{V\cap W}(x_i) = \min\{\mu_V(x_i), \mu_W(x_i)\} \tag{10}$$

$$V \subset W \text{ iff } \mu_V(x_i) \leq \mu_W(x_i), \forall x_i \in U \tag{11}$$

$$\text{supp } V = \{x_i \in U; \mu_V(x_i) > 0\} \tag{12}$$

$$\text{card } V = \sum_i \mu_V(x_i) \tag{13}$$

Consider the fuzzy partitioning of a universe of discourse U by the attributes in B, defined as U/B. The positive region in classical rough set is defined as the union of lower approximations [18]. Applying the extension principle, the membership of an object $x_i \in U$ belonging to the fuzzy positive region can be defined as

$$\mu_{\mathrm{POS}_{\tilde{R}Y}}(x_i) = \sup_{x_i\in U/B} \mu_{\underline{\tilde{R}}Y}(x_i). \tag{14}$$

The fuzzy rough dependency function may be defined as [18] using definition of fuzzy positive region as follows:

$$\gamma_{\tilde{R}Y} = \frac{\left|\mu_{\mathrm{POS}_{\tilde{R}Y}}(x_i)\right|}{|U|} = \frac{\sum_{x_i\in U} \mu_{\mathrm{POS}_{\tilde{R}Y}}(x_i)}{|U|} \tag{15}$$

It can be observed that FRS and the related concepts are based on the membership function. In the following sections, we develop the linguistic rough set.

3 Linguistic approximation space

Here, we give the motivations for the linguistic rough set, followed by the formulation of linguistic approximation space.

3.1 Motivation

In many decision making situations, often it is difficult for DMs to assess an alternative quantitatively in terms of a crisp number or even assign a fuzzy membership value. At best, the facts are represented by natural language [1–14]. For example: I wish to buy a car which should be low cost, high mileage and medium green in color. In order to process such information, the conventional fuzzy theory advocates constructing fuzzy sets to represent each of the quantifiers like “low”, “very high”, and “medium green”, and then determine the membership degree in terms of a fuzzy number to indicate the quality of the information. However, in many decision making situations, a decision maker (DM) has only partial knowledge about this membership degree. In other words, she is having a broad range in mind for a quantifier, say “medium green”. Casting this imprecision through MF sometimes distort the original intended imprecision expressed in the statements.

The linguistic quantifiers such as “medium green” also play the role of fuzzy constraints. From this perspective, the linguistic quantifiers may be regarded as a form of fuzzy information granules that aid in representation of imprecision through information summarization. The close relationship between linguistic quantifiers and fuzzy information granulation is discussed in [12]. The information granulation provides the strong fulcrum on which the concept of linguistic quantifier rests. At the same time, information granulation forms an integral part of human cognition, thought process, information processing and decision making. The fuzziness of granules is characteristic of ways in which humans granulate and manipulate information [1, 12]. Granulation is also universally observed in nature and the real world around us. Any matter can be viewed to be comprising of fine granules of similar type, composition of which forms the matter in question. Zadeh in [12] identified granulation, along with organization and causation, as one of primary concepts underlying human cognition. Therefore information granulation has to be the core of any intelligent system or methodology for representation of the real world imprecision.

Information granulation caters to partial knowledge, partial understanding, partial belief and partial certainty in the real world. We often classify the objects together that look broadly similar to us for information simplification. For example, all domestic cats, lions and tigers belong to Cat or Felidae family because of the common traits like muscular body, twisted front legs bones, soft pads in paws etc. In view of the available knowledge of traits they appear indiscernible and hence are classified under one family. With access to more information, they can be further classified into subfamilies: Pantherinae (that includes tigers and lions) and Felinae (that includes domestic cats). Further at a macroscopic level they all fall into order Carnivore and class Mammalia. Each class or these categories can be viewed as a granule of similar or indiscernible objects in view of the available information. A linguistic quantifier with underlying fuzzy granulation offers a good potential in modelling of such natural representations [23, 24].

The rough set framework has been widely accepted as an approximation tool with incomplete knowledge. In the rough set theory, any set of all indiscernible objects is called an elementary set or an information granule. A collection of such information granules may also be referred to as elementary knowledge. In the conventional rough set and its existing generalizations, this elementary knowledge is either in the form of crisp values or MF values. Combining linguistic quantifiers with the rough set theory replaces the numerical (or MF) based elementary knowledge with words that are the labels of knowledge granules. The resulting elementary knowledge that forms the grammar for approximation of imprecise concepts is termed as linguistic approximation space (LAS). The common principle of granulation shared by fuzzy quantifiers and rough set framework facilitate the generation of LAS.

3.2 Linguistic approximation space

There lies a clear symmetry between the rough set and linguistic quantifier frameworks in the sense that both are characterized by indiscernible objects or granules of knowledge about the universe. This symmetry aids the integration of these two uncertainty representation formalisms. The proposed linguistic approximation space (LAS) is characterized with the linguistic hedges such as very low, low, high etc., which are synonymous with an information granule. The objects that are similar, or are in close proximity to each other, are assigned the same linguistic quantifier and are classified under the same granule. LAS operates directly upon these linguistic quantifiers, which greatly simplifies the task of dealing with large and complex real valued data sets in real life applications by achieving data compression and summarization of data. All approximations are in terms of these linguistic quantifiers in a process analogous to the granulate-and-simplify strategy, commonly used by the human brain.

LAS is characterized with a finite and totally ordered discrete linguistic label set $S = \{s_i | i = -z, \ldots, -1, 0, 1, \ldots, z\}$, where z is a positive integer, s_i represents a linguistic label [25], and satisfies $s_i > s_j$ if $i > j$. An example of a discrete linguistic label set is shown as follows:

$$\begin{aligned} S_1 = \{ & s_{-4} = \text{extremely poor}, s_{-3} = \text{very\ \ poor}, \\ & s_{-2} = \text{poor}, s_{-1} = \text{slightly\ \ poor}, s_0 = \text{medium}, \\ & s_0 = \text{slightly\ \ good}, s_2 = \text{good}, \\ & s_3 = \text{very\ \ good}, s_4 = \text{extremely good}\}. \end{aligned}$$

The linguistic set S_1 can be used to express the assessments of alternatives. This method is especially useful to express human perceptions or judgments with partial knowledge. The use of numbers in such cases brings in unwanted precision that distorts the original imprecise assessment. This concern remains unaddressed with the fuzzy set theory as the representation of uncertainty through membership grades lead to the loss of original vagueness inherent in the linguistic representations. The direct use of linguistic quantifiers suitably addresses this concern.

The linguistic quantifiers in S also caters to partial knowledge and understanding of an observer, or underlying vagueness in the real world. So when two alternatives P and Q are assessed by an expert as good and poor respectively against a criterion, the expert has a wide range in his mind for "good" and "poor". The alternatives P and Q fit well within this range and hence they are respectively assigned these linguistic labels. Here, we again emphasize that "good" and "poor" are the easily interpretable terms chosen from the natural language, and they label a set (granule) of objects, falling under a category depending upon their similarity. The use of such linguistic quantifiers let an expert to distinguish between two alternatives without undesired precision creeping in distorting the original intent. The linguistic quantifiers are imprecise in their basic character, yet precise enough to compare and evaluate different alternatives.

Therefore a delicate balance needs to be maintained in choosing the cardinality of set S. If the quantifiers are too fine-tuned, it would lead to the loss of the original underlying imprecision. Similarly, broadly spaced linguistic quantifiers would bring in too much generalization at the cost of accuracy of approximation. For example, consider a set of linguistic labels S_2 as shown

$S_2 = \{s_{-1} = \text{poor}, s_0 = \text{medium}, s_1 = \text{good}\}.$

An expert using such a system to evaluate alternatives would be seriously incapacitated due to lack of choices he is having at his disposal to grade the alternatives.

In comparison, let us consider another set S_3 as shown

$$\begin{aligned} S_3 = \{ & s_{-8} = \text{pathetic},\ s_{-7} = \text{deplorable}, \\ & s_{-6} = \text{miserable},\ s_{-5} = \text{feeble},\ s_{-4} = \text{extremely poor}, \\ & s_{-3} = \text{very poor}, s_{-2} = \text{poor}, s_{-1} = \text{slightly poor}, \\ & s_0 = \text{medium}, s_1 = \text{slightly good}, s_2 = \text{good}, \\ & s_3 = \text{very good}, s_4 = \text{extremely good}, s_5 = \text{excellent}, \\ & s_6 = \text{outstanding}, s_7 = \text{exquisite}, s_8 = \text{elegant}\}. \end{aligned}$$

Such a set is fine-tuned with more number of linguistic quantifiers. In summary, the choice of linguistic quantifiers should be directly in line with the context of the problem. If the number of alternatives to be compared is very large and expert is confidently able to give his assessments over the finely grained linguistic quantifiers, S_3 might be an ideal choice. On the other hand, if only a few alternatives are to be compared or there are wide differences in the performance of alternatives, S_1 or S_2 should be fine. The formal definitions related to LAS now follow.

Definition 3.1 A linguistic approximation space can be defined as a system $(U, C, \mathcal{L})$, where $U = \{x_1, x_2, \ldots, x_m\}$ and C is the set of attributes (features, variables). The family of attributes is also known as knowledge in the universe. Each attribute $c \in C$ defines an information function $f_c : U \to \mathcal{L}$ where $\mathcal{L}$ is the set of linguistic labels for the approximation space. A granular approximation space may also be termed as granular information system (GIS).

$$\text{GIS} = (U, C, \mathcal{L}) \tag{16}$$

Example 3.1 An oil mining expert performs the assessment of five oil-fields $\{x_1, x_2, \ldots, x_5\}$ for mining oil in terms of three criteria: geological structure (C_G), quality of oil (C_q), and legal position (C_l). The results of the assessment are given as follows.

In Table 1, five alternatives $(U = \{x_1, x_2, \ldots, x_5\})$ are described by three attributes or criteria $(C = \{C_G, C_q, C_l\})$. Each alternative is assigned a linguistic quantifier from the set S_1 (Here, $\mathcal{L} = S_1$) as its assessment against a criterion. Each of these linguistic quantifiers is analogous with a knowledge granule.

Example 3.2 Consider the information system of Table 1. Determine the indiscernible alternatives in terms of linguistic quantifiers.

In this example, x_1 and x_3 form one equivalence class over the attribute C_G as both the alternatives have the same linguistic value, slightly poor (s_{-1}) against the criterion of geological structure (C_G). Similarly, the alternatives x_1 and x_4 are indiscernible over LAS $(U, C, \mathcal{L})$, with $C = (C_q, C_l)$ and $\mathcal{L} = S_1$.

4 Linguistic rough sets

This section is devoted to the concept of linguistic rough set. We first discuss the concept of inclusion degree that is central to linguistic rough set, followed by a presentation of the linguistic rough set.

4.1 Inclusion degree

The concept of inclusion/covering degree is introduced into rough set theory in [26, 27]. Recently, some properties of covering-based rough set are investigated in [39]. Since the basic motivation behind LRS is to retain the underlying imprecision, robustness has to be at the heart of the approximations obtained through LRS, so that they could withstand the inaccuracies in uncertainty assessments. With this goal in sight, we have used the concept of inclusion degree as the basis for LRS, which is in stark contrast to the fuzzy equivalence relation used in FRS. Before we delve upon the concept of linguistic rough set, we adapt the concept of inclusion degree based on the partially ordered relation in the context of LAS.

Table 1 Simple granular information system

Criteria	C_G	C_q	C_l
Alternatives			
x_1	s_{-1}	s_4	s_{-3}
x_2	s_0	s_1	s_3
x_3	s_{-1}	s_{-4}	s_{-1}
x_4	s_2	s_4	s_{-3}
x_5	s_1	s_2	s_{-2}

A partially ordered set $(L, \preccurlyeq)$ is a binary relation $\preccurlyeq$ with the following properties of reflexivity, antisymmetry and transitivity [26]:

1. $x \preccurlyeq x$
2. $x \preccurlyeq y$ and $y \preccurlyeq x \implies x = y$ and,
3. $x \preccurlyeq y$ and $y \preccurlyeq z \implies x \preccurlyeq z$. (17)

Let $(L, \preccurlyeq)$ be a partially ordered set. For any $(V, W) \in L$, the inclusion degree, $\mathcal{D}$ on L is defined as a real number $\mathcal{D}(V, W)$ with the following properties:

1. $0 \leq \mathcal{D}(V, W) \leq 1$,
2. $V \preccurlyeq W \Longrightarrow \mathcal{D}(W, V) = 1$,
3. $V \preccurlyeq W \preccurlyeq Z \Longrightarrow \mathcal{D}(W, V) \leq \mathcal{D}(Z, V)$. (18)

We now give the definition of inclusion degree in the context of linguistic approximation space.

Definition 4.1 Consider a LAS $(U, C, \mathcal{L})$ over the non-empty and finite set of objects U. We recall that $U = \{x_1, x_2, \ldots.., x_m\}$. Let $C = \{C_1, C_2, \ldots \ldots, C_n\}$ be a family of concepts from the power set of U, denoted as $\mathcal{F}(U)$, which forms a partition of U; $\mathcal{D}(W, V)$ be the degree of inclusion of concept V in concept W; $V(x_i)$ and $W(x_i)$ be the corresponding linguistic quantifiers for an object $x_i \in U$; and $\mathcal{I}(W, V) = \{V(x_i) \leq W(x_i)\}, \forall x_i \in U$. Then the degree of inclusion $\mathcal{D}(W, V)$ for $W, V \in \mathcal{F}(U)$ for LAS $(U, C, \mathcal{L})$ is defined as

$$\mathcal{D}(W, V) = \frac{|\mathcal{I}(W, V)|}{|V|} \tag{19}$$

where, $|.|$ denotes the cardinality of a set or the count of elements. Note: If $V = \phi$ then $\mathcal{D}(W, V) = 1$.

The inclusion degree $\mathcal{D}(W, V)$ gives the relative count of objects for which $V(x_i) \leq W(x_i)$. If this condition is true for all objects $x_i \in U$, then concept V is fully included in W. Since, D depends on linguistic quantifiers, it is not sensitive to membership degrees. In comparison, FRS is based on membership grades that are assumed to be sacrosanct. Any inaccuracy in membership grade directly affects the approximation arrived at through FRS. These kinds of errors are drastically minimized in the proposed approach due to the following facts

- The linguistic quantifiers are synonymous with wide knowledge granules, which obviate the need to assign numerical membership grades. As a result, the approximations obtained are generalized and broad-based.
- Since the inclusion degree is the relative count of objects, it is not much affected with a noisy sample. For instance, with the presence of a noisy sample, the inclusion degree is not affected as far as the condition $V(x_i) \leq W(x_i)$ remains true. Even in the case of this condition getting violated (falsely) due to erroneous $V(x_i)$ or $W(x_i)$, the impact is only $^1/_{|V|}$ on $\mathcal{D}(W, V)$.

Another advantage that is imparted to LRS with inclusion degree as its basis is the generalization in the approximations. In FRS, the approximations are rigidly true only with respect to the given LAS, and not for the larger population (due to deviations from the given LAS). In contrast, the controllable inclusion degree parameter in LRS helps to control the trade-off between generalization and accuracy of the approximations. A balance could be maintained between the two by choosing $\mathcal{D} \in [0, 1]$ as per the context and desirability. Full covering or complete inclusion of concept V by W is achieved when $\mathcal{D}(W, V) = 1$, which guarantee maximum accuracy (of approximations holding good with respect to the given LAS) but nil generalization.

We now give yet another measure of accuracy (or generalization) of the approximation obtained through LRS. We term this measure as degree of certainty of approximations, and it is based on the concept of inclusion degree.

Definition 4.2 We consider the family of concepts $C = \{C_1, C_2, \ldots \ldots, C_n\}$ from $\mathcal{P}(U)$ and $Y \in \mathcal{P}(U)$. The lower and upper bounds for the degree of certainty of approximations are denoted by λ_L and λ_U, and are given as

$$\lambda_L = \mathcal{D}\left(Y, \bigcap_{C_j \in C} C_j\right)$$
$$\lambda_U = \mathcal{D}\left(\bigcup_{C_j \in C} C_j, Y\right) \tag{20}$$

The net degree of certainty λ is given as

$$\lambda = \min\{\lambda_L, \lambda_U\} \tag{21}$$

where, $\bigcap_{C_j \in C} C_j = \min_j \{C_j(x_i)\}_{i=1}^m$ and $\bigcup_{C_j \in C} C_j = \max_j \{C_j(x_i)\}_{i=1}^m$. λ_L indicates the level of certainty that $\cap C_j(x_i) \leq Y(x_i)$ for $\forall x_i \in U$ and λ_U indicates the level of certainty that $Y(x_i) \leq \cup C_j(x_i)$ for $\forall x_i \in U$. Also, $\lambda_L, \lambda_U, \lambda \in [0, 1]$. If $\lambda_L = \lambda_U = 1$, it can be concluded that $\cap C_j(x_i) \leq Y(x_i) \leq \cup C_j(x_i), \forall x_i$. If the approximations of Y are obtained with degree of certainty λ, then we say Y is λ-approximable.

The degree of certainty $\lambda \in [0, 1]$ indicates the level of generalization or accuracy in the approximation. The nearer it is to 1, more accurate is the approximation obtained with respect to the given LAS, however lesser is its generalization for a larger population with deviations from the values in the given LAS. The choice of λ_L and λ_U

provides decision makers a convenient lever to control the trade-off between the generalization and accuracy levels.

4.2 Linguistic rough sets

We now present the concept of linguistic rough set to approximate a decision concept Y in terms of linguistic quantifiers assigned to various objects against condition concepts set C.

Definition 4.3 Let $Y \in \mathcal{P}(U)$ be a λ-approximable set in LAS $\langle U, C, \mathcal{L} \rangle$; *P(Y)* denote the set $\{K \subset C; \mathcal{D}(Y, \cap_{C_j \in K} C_j) \geq \lambda\}$; and *Q(Y)* denote the set $\{L \subset C; \mathcal{D}(\cup_{C_j \in L} C_j, Y) \geq \lambda\}$. Let $K^* \in P(Y)$ and $L^* \in Q(Y)$ be such that

$$\text{card}\left(\underset{C_j \in L^*}{\cup} C_j - \underset{C_j \in K^*}{\cap} C_j\right) = \min_{K \in P(Y), L \in Q(Y)} \text{card}\left(\underset{C_j \in L}{\cup} C_j - \underset{C_j \in K}{\cap} C_j\right). \tag{22}$$

Then the lower and upper approximations of Y in LAS $\langle U, C, \mathcal{L} \rangle$ are given by

$$\underline{C}_\lambda(Y) = \underset{C_j \in K^*}{\cap} C_j$$

$$\overline{C}_\lambda(Y) = \underset{C_j \in L^*}{\cup} C_j. \tag{23}$$

The pair $(\underline{C}_\lambda(Y), \overline{C}_\lambda(Y))$ is called linguistic rough set. It is possible to obtain more than one pair of approximation of Y with the proposed linguistic rough set in case there are multiple pairs of sets K^* and L^* such that $K^* \in P(Y)$ and $L^* \in Q(Y)$. It can also be observed that the identification of the attribute sets K and L ensures a balance between quality of approximation and generalization. The condition in (22) is meant to ensure crispness in approximations. We now give separate measures for quality and crispness in approximations.

Definition 4.4 The quality of approximation for linguistic rough set is indicated by tolerance coefficient and is given by

$$T_C(Y) = \frac{\left|\overline{C}_\lambda(Y) - \underline{C}_\lambda(Y)\right|}{|U|}. \tag{24}$$

Definition 4.5 The crispness coefficient for the approximations obtained with LRS is obtained by taking the difference between 1 and the ratio of sum of differences of lower and upper approximations of various objects and the count of objects in U. That is

$$\kappa = 1 - \frac{\sum_i \left(\cup_{C_j \in L^*} C_j(x_i) - \cap_{C_j \in K^*} C_j(x_i)\right)}{|U|}. \tag{25}$$

Definition 4.6 The accuracy of approximation is measured by the coefficient

$$\alpha_\lambda(Y) = \frac{\left|\underline{C}_\lambda(Y)\right|}{\left|\overline{C}_\lambda(Y)\right|}. \tag{26}$$

We now discuss some of the properties of concepts that hold true in a linguistic approximation space.

Theorem 4.1 The following properties stand true for any two concepts $V, W \in \mathcal{P}(U)$ in LAS $\langle U, C, \mathcal{L} \rangle$.

The upper approximation of a set of condition attributes is the set itself. That is

$\overline{C}_\lambda(C) = C.$

The upper approximation of an empty set is the empty set. That is

$\overline{C}_\lambda(\phi) = \phi.$

The properties related to lower and upper approximations of union or intersection of two concepts are given as follows.

$$\underline{C}_\lambda(V \cap W) = \underline{C}_\lambda(V) \cap \underline{C}_\lambda(W)$$

$$\overline{C}_\lambda(V \cup W) = \overline{C}_\lambda(V) \cup \overline{C}_\lambda(W)$$

$$\underline{C}_\lambda(V) \cup \underline{C}_\lambda(W) \subseteq \underline{C}_\lambda(V \cup W)$$

$$\overline{C}_\lambda(V \cap W) \subseteq \overline{C}_\lambda(V) \cap \overline{C}_\lambda(W).$$

The following properties show that if V is a subset of W, then the lower and upper approximations of V are bound to be subsets of that of W.

$$V \subseteq W \implies \underline{C}_\lambda(V) \subseteq \underline{C}_\lambda(W)$$

$$V \subseteq W \implies \overline{C}_\lambda(V) \subseteq \overline{C}_\lambda(W).$$

Proof The proof follows directly from (23).

The above properties lead us to the following properties that are about intersection and union of a concept with itself.

Theorem 4.2 Given a LAS $\langle U, C, \mathcal{L} \rangle$ and concepts $V, W \in \mathcal{P}(U)$, we have

1. $\overline{C}_\lambda(V) = \overline{C}_\lambda(W) \iff \overline{C}_\lambda(V) = \overline{C}_\lambda(V \cup W) = \overline{C}_\lambda(W);$
2. $\underline{C}_\lambda(V) = \underline{C}_\lambda(W) \iff \underline{C}_\lambda(V) = \underline{C}_\lambda(V \cap W) = \underline{C}_\lambda(W)$
3. $\overline{C}_\lambda(V) = \overline{C}_\lambda(W)$ and $\underline{C}_\lambda(V) = \underline{C}_\lambda(W) \implies V = W$

Proof

1. We are given that $\overline{C}_\lambda(V) = \overline{C}_\lambda(W)$. Also, from Theorem 3.1, we have $\overline{C}_\lambda(V \cup W) = \overline{C}_\lambda(V) \cup \overline{C}_\lambda(W)$, so we have $\overline{C}_\lambda(V \cup W) = \overline{C}_\lambda(V) = \overline{C}_\lambda(W)$. Therefore, $\overline{C}_\lambda(V) = \overline{C}_\lambda(W) \iff \overline{C}_\lambda(V) = \overline{C}_\lambda(V \cup W) = \overline{C}_\lambda(W)$.

Conversely, if $\overline{C}_{\lambda}(V) = \overline{C}_{\lambda}(V \cup W) = \overline{C}_{\lambda}(W)$ then by transitivity we have $\overline{C}_{\lambda}(V) = \overline{C}_{\lambda}(W)$.

2. Similarly, if $\underline{C}_{\lambda}(V) = \underline{C}_{\lambda}(W)$, then $\underline{C}_{\lambda}(V) = \underline{C}_{\lambda}(V \cap W) = \underline{C}_{\lambda}(W)$. The converse is also true by transitivity.
3. The proof straightforwardly follows from (23).

The above properties show that the approximations in LRS follow the general properties of sets. This adds to the usefulness of LRS. On the similar lines, we now prove a few more properties of LRS.

Theorem 4.3 Let us consider concepts $V, W, V_1, W_1 \in \mathcal{P}(U)$ in LAS $\langle U, C, \mathcal{L} \rangle$, then we have

1. If $\overline{C}_{\lambda}(V) = \overline{C}_{\lambda}(V_1), \overline{C}_{\lambda}(W) = \overline{C}_{\lambda}(W_1) \implies \overline{C}_{\lambda}(V \cup W) = \overline{C}_{\lambda}(V_1 \cup W_1)$;
2. If $\underline{C}_{\lambda}(V) = \underline{C}_{\lambda}(V_1), \underline{C}_{\lambda}(W) = \underline{C}_{\lambda}(W_1) \implies \underline{C}_{\lambda}(V \cap W) = \underline{C}_{\lambda}(V_1 \cap W_1)$

Proof

1. We are given $\overline{C}_{\lambda}(V) = \overline{C}_{\lambda}(V_1)$ and $\overline{C}_{\lambda}(W) = \overline{C}_{\lambda}(W_1)$. From Theorem 3.1, we know that $\overline{C}_{\lambda}(V \cup W) = \overline{C}_{\lambda}(V) \cup \overline{C}_{\lambda}(W)$. Therefore, $\overline{C}_{\lambda}(V \cup W) = \overline{C}_{\lambda}(V_1) \cup \overline{C}_{\lambda}(W_1) = \overline{C}_{\lambda}(V_1 \cup W_1)$. That is $\overline{C}_{\lambda}(V \cup W) = \overline{C}_{\lambda}(V_1 \cup W_1)$..
2. From Theorem 3.2, we have $\underline{C}_{\lambda}(V \cap W) = \underline{C}_{\lambda}(V) \cap \underline{C}_{\lambda}(W)$. Therefore, $\underline{C}_{\lambda}(V \cap W) = \underline{C}_{\lambda}(V_1) \cap \underline{C}_{\lambda}(W_1) = \underline{C}_{\lambda}(V_1 \cap W_1)$. That is $\underline{C}_{\lambda}(V \cap W) = \underline{C}_{\lambda}(V_1 \cap W_1)$.

Similarly, the following two properties of LRS are also useful in decision making applications.

Theorem 4.4 For any two concepts $V, W \in \mathcal{P}(U)$ in LAS $\langle U, C, \mathcal{L} \rangle$, the following relations stand true.

1. If $V \subseteq W$, $\overline{C}_{\lambda}(W) = \emptyset \implies \overline{C}_{\lambda}(V) = \emptyset$;
2. If $V \subseteq W$, $\underline{C}_{\lambda}(V) = C_j \implies \underline{C}_{\lambda}(W) = C_j$

Proof

1. We have $V \subseteq W$, $\underset{\lambda}{\overline{C}}(W) = \emptyset$, and from Theorem 3.1, we know that $\overline{C}_{\lambda}(V \subseteq W) = \overline{C}_{\lambda}(V) \subseteq \emptyset$. This implies that $\overline{C}_{\lambda}(V) = \emptyset$.
2. From Theorem 3.1, we know that $\underline{C}_{\lambda}(V \subseteq W) = \underline{C}_{\lambda}(V) \subseteq \underline{C}_{\lambda}(W)$. Since, $\underline{C}_{\lambda}(V) = C_j$, $\underline{C}_{\lambda}(W) = C_j$.

5 Application of linguistic rough set in decision making

In this section we show an application of the proposed linguistic rough set in a real world case-study.

5.1 Approximation of fuzzy decision concept with linguistic rough set

We consider a decision making framework as given in Fig. 1. Here, $U = \{x_i\}, i = 1, \ldots, m$ represents a set of alternatives. A set $C = \{C_j\}, j = 1, \ldots, n$ denotes the set of fuzzy concepts, in terms of which a given concept should be approximated. These concepts are referred to as condition concepts, in the sequel. We consider that the experts only have partial information to evaluate an alternative from U against each of the condition concepts $\{C_j\}, j = 1, \ldots, n$. It is easier for the experts to assess an alternative broadly in terms of the linguistic quantifiers. Let c_{ij} denote the assessment (in terms of a linguistic quantifier) of alternative x_i against condition concept C_j.

Our objective is to approximate a set of fuzzy concepts, denoted by $Y = \{Y_1, Y_2, \ldots, Y_t\}$, in terms of condition concepts $\{C_j\}, j = 1, \ldots, n$, on the basis of linguistic assessments $\{c_{ij}\}; i = 1, \ldots, m; j = 1, \ldots, n$. We refer to the concepts in set Y as decision concepts.We now perform the following steps and apply the proposed linguistic rough set to approximate a decision concept.

Step 1 We construct an information system with the given alternatives, say $U = \{x_1, x_2, \ldots, x_m\}$, condition concepts $C = \{C_1, C_2, \ldots, C_n\}$ and decision concepts $Y = \{Y_1, Y_2, \ldots, Y_t\}$. An alternative x_i is assessed against the various condition concepts from C in terms of the linguistic labels from linguistic set $\mathcal{L}$. The weights of various condition concepts may be taken into account while arriving at the assessments for decision concepts Y.

The information system may be represented by a matrix, $[I]_{m \times n}$. Each entry (i_{ij}) in the matrix $[I]_{m \times n}$ is a linguistic term denoting the assessment of ith alternative against jth criterion. A sample matrix $[I]_{m \times n}$ is shown in Table 2.

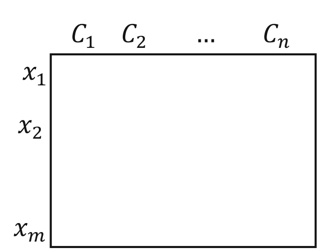

Fig. 1 Multi criteria decision making paradigm

Table 2 Granular information system $[I]_{m\times n}$

Alternatives	C_1	C_2	…	C_j	…	C_n	Y_1	Y_2	…	Y_k	…	Y_t
Alternatives												
x_1	c_{11}	c_{12}	…	c_{1j}	…	c_{1n}	y_{11}	y_{12}	…	y_{1k}	…	y_{1t}
x_2	c_{21}	c_{22}	…	c_{2j}	…	c_{2n}	y_{21}	y_{22}	…	y_{2k}	…	y_{2t}
…	…	…	…	…	…	…	…	…	…	…	…	…
x_i	c_{i1}	c_{i2}	…	c_{ij}	…	c_{in}	y_{i1}	y_{i2}	…	y_{ik}	…	y_{it}
…	…	…	…	…	…	…	…	…	…	…	…	…
x_m	c_{m1}	c_{m2}	…	c_{mj}	…	c_{mn}	y_{m1}	y_{m2}	…	i_{mk}	…	i_{mt}

m is the total number of alternatives, n is the total number of condition concepts, and t is the total number of decision concepts

Step 2 We find the degree of inclusion of each condition concept C_j in each decision concept Y_k by applying (19). The degree of inclusion of jth conditional concept in kth decision concept, $\mathcal{D}(Y_k, C_j)$ is computed by applying (19) as

$$\mathcal{D}(Y_k, C_j) = \frac{\left|\left\{c_{ij} \le y_{ik}; i = 1, 2, \ldots m\right\}\right|}{m}.$$

Step 3 Similarly, we also find the degree of inclusion of each decision concept in each condition concept by applying (19). The degree of inclusion of kth decision concept in jth condition concept, $\mathcal{D}(C_j, Y_k)$ is found as

$$\mathcal{D}(C_j, Y_k) = \frac{\left|\left\{y_{ik} \le c_{ij}; i = 1, 2, \ldots m\right\}\right|}{m}.$$

Step 4 To compute the lower approximation of Y_k, we look for highest degree of inclusion/covering of any C_j by Y_k. The degree of covering indicates certainty in the approximation. A degree of covering $\mathcal{D}(Y_k, C_j)$ of 1 reflects that the C_j is fully covered by Y_k. We look for the highest degree of covering so as to ensure the maximum possible certainty in the lower approximation of a decision concept. That is, we look for C_j such that

$$C_j : \max\left(\mathcal{D}(\mathrm{Y_k}, \mathrm{C_j})\right); j = 1, 2, \ldots, n.$$

The condition concept C_j corresponding to $\mathcal{D}(\mathrm{Y_k}, \mathrm{C_j}) = 1$ would be the lower approximation for decision concept $\mathrm{Y_k}$.

Step 5 If the value of $\mathcal{D}(Y_k, C_j) < 1$, we compute the degree of inclusion of C_j in combination with other condition concepts by Y_k in an attempt to obtain a higher degree of covering by Y_k. That is

$$C_j \cap C_u : \max\left(\mathcal{D}\left(Y_k, C_j \cap C_u\right)\right); u = 1, 2, \ldots, n; u \ne j.$$

Step 6 In case the value of $\left(\mathcal{D}\left(Y_k, C_j \cap C_u\right)\right) < 1$, we repeat Step 4 with combination of $C_j \cap C_u$ and remaining condition concepts in a bid to obtain a higher degree of inclusion than the one obtained so far, until a degree of inclusion of 1 is obtained or all the condition concepts are exhausted. That is

$$C_j \cap C_u \cap \ldots : \max\left(\mathcal{D}\left(Y_k, C_j \cap C_u \cap \ldots\right)\right).$$

Step 7 We compute the upper approximation of Y_k by looking for highest degree of inclusion of Y_k by any C_j. A degree of covering $\mathcal{D}(C_j, Y_k)$ of 1 reflects that Y_k is fully covered by jth condition concept

$$C_{j'} : \max\left(\mathcal{D}(C_j, Y_k)\right); j = 1, 2, \ldots, n.$$

The condition concept C_j corresponding to $\mathcal{D}\left(\mathrm{C_j}, \mathrm{Y_k}\right) = 1$ would be the upper approximation for decision concept Y_k.

Step 8 If the value of $\mathcal{D}\left(C_{j'}, Y_k\right) < 1$, we compute the degree of inclusion of Y_k by $C_{j'}$ in union with other condition concepts to check for the possibility of obtaining a higher degree of inclusion of Y_k by the union of condition concepts. That is

$$C_{j'} \cup C_u : \max\left(\mathcal{D}\left(Y_k, C_{j'} \cup C_u\right)\right); u = 1, 2, \ldots, n; u \ne j'.$$

Step 9 If the value of $\mathcal{D}\left(C_{j'} \cup C_u, Y_k\right) < 1$, we repeat Step 8 with union of $C_{j'} \cup C_u$ and remaining condition concepts in a bid to obtain a higher degree of inclusion than the one obtained so far until a degree of inclusion of 1 is obtained or all the condition concepts are exhausted. That is

$$C_{j'} \cup C_u \cup \ldots : \max\left(\mathcal{D}\left(Y_k, C_{j'} \cup C_u \cup \ldots\right)\right).$$

Step 10 We repeat Steps 4–10 for approximating each of the decision concepts.

5.2 Case illustration

We now apply the approach, conceived above, in an oil investment case [28] characterized by linguistic assessments by experts. It may be intuitively felt that when vague or linguistic assessments are involved, the decisions in such situations should also not be crisp but little vague so as to represent a wide range of values in commensuration with linguistic inputs. The approximations obtained through LRS represent a wide range of acceptable values that fit well with the linguistic nature of the decision making problem.

The suitability of a region as a potential oil investment case depends broadly upon the perception of prospectivity (C_1), exploration programme (C_2), favorable legal position (C_3), favorable fiscal position (C_4), stratification position (C_5) and potential of the area for supporting oil discovery (C_6). It can be easily observed that given these condition concepts, the experts may not be completely knowledgeable to arrive at the precise evaluations of alternatives (objects) against these condition concepts in terms of crisp numbers. In such situations, linguistic quantifiers greatly simplify the task of evaluation of alternatives. At the same time, the rough set framework helps to approximate imprecise decision concepts. Hence, the proposed linguistic rough set is of special use in such applications.

The weight vector of the condition concepts is $w = (0.25, 0.18, 0.22, 0.15, 0.09, 0.11)$. The actions of the investment may lead to three possible results: Invest (Y_1), Not invest (Y_2) and Defer (Y_3), which form the decision concepts to be approximated. The decision makers (DMs) evaluate the prospective regions (alternatives) $x_i(i = 1,\ldots 8)$ with respect to the criteria $C_j(j = 1,\ldots,6)$ and assign linguistic terms from the set S that is shown as

$$S = \{s_{-4} = \text{extremely poor}, s_{-3} = \text{very poor}, s_{-2} = \text{poor}, s_{-1} = \text{slightly poor}, s_0 = \text{medium}, s_1 = \text{slightly good}, s_2 = \text{good}, s_3 = \text{very good}, s_4 = \text{extremely good}\}.$$

The DMs assign a linguistic term from S against each of the decision concepts $Y_i(i = 1, 2, 3)$. The linguistic information system so generated is shown in Table 3. We now approximate of each of the decision concepts Y_1, Y_2 and Y_3 in terms of the condition concepts by applying proposed linguistic rough set framework. The approximations are arrived at on the basis of the linguistic information system, shown in Table 3.

Solution We now perform the steps as outlined in Sect. 5.1 to approximate Y_1, Y_2 and Y_3.

Step 1 We generate the linguistic information system as given in Table 3 with condition concepts C_j ($j = 1,\ldots, 6$), decision concepts Y_i ($i = 1, 2, 3$) and universe of discourse represented by the set of alternatives $x_i(i = 1,\ldots,8)$.

Step 2 The degree of inclusion of C_j in Y_k, denoted by $\mathcal{D}(Y_k, C_j)$, is computed by applying (19). The values $\mathcal{D}(\mathrm{Y_k}, \mathrm{C_j}); j = 1,\ldots,6; k = 1,\ldots,3$ are reported in Table 4.

Step 3 Similarly, we compute the degree of inclusion of Y_k in C_j, denoted by $\mathcal{D}(C_j, Y_k)$, and populate the values in Table 5.

The degrees of inclusion in Tables 4 and 5 form the basis of approximation of decision concepts $Y_k, k = 1,\ldots,3$. We now find the lower and upper approximations of $Y_k, k = 1,\ldots,3$ by applying the proposed linguistic rough set.

Approximation of Y_1

Step 4 We find the lower approximation of decision concept Y_1. From Table 4, C_3 is seen to be having the highest degree of covering by Y_1 at 0.625. Since this value is less than 1, we look for other conditional concepts in combination of which we could obtain a higher degree of inclusion, and hence a higher degree of certainty in lower approximation of Y_1.

Step 5 We find C_j such that $\mathcal{D}(Y_1, C_3 \cap C_j) > 0.625; j = 1,\ldots,6; j \neq 3$. The process is continued until all the conditional concepts C_j are exhausted or the degree of inclusion of 1 is obtained.

Step 6 The combination $C_1 \cap C_2 \cap C_3$ is found to be having the degree of inclusion as 1, that is $\mathcal{D}(Y_1, C_1 \cap C_2 \cap C_3) = 1$. Here, it is recalled that a degree of inclusion of 1 indicates that the lower approximation of Y_1 is $C_1 \cap C_2 \cap C_3$ with a degree of certainty of 100 %.

Step 7 The upper approximation of decision concept Y_1 is found in this step. From Table 5, the highest degree of inclusion of Y_1 by any C_j is 1 for C_2, C_5 and C_6.

It is found that an alternative x_i belongs to the granule "Invest" with the linguistic coefficient of membership at least as high as minimum of its membership in C_1, C_2 and C_3. Also the propensity to invest is bounded by an upper limit of the maximum linguistic value among the assessment of x_i against the condition concepts C_2, C_5 and C_6. In other words, the linguistic assessment of an alternative in the granule "invest" would lie somewhere between the lower limit of $C_1 \cap C_2 \cap C_3$ and the upper limit of $C_2 \cup C_5 \cup C_6$.

Approximation of Y_2

It can be seen from Table 4 that the highest degree of covering of any C_j by Y_2 is 0.75 for C_3. Therefore, we would find C_j such that $\mathcal{D}(Y_2, C_3 \cap C_j) = 1, j = 1,\ldots,6; j \neq 3$. No such combination is found, which means with the given data, it is not possible to have a degree of

Table 3 Linguistic information system for potential regions of oil investment

x_i	Prospectivity C_1	Exploration program C_2	Legal position C_3	Fiscal position C_4	Stratification position C_5	Potential of area C_6	Y_1	Y_2	Y_3
1	s_0	s_4	s_{-2}	s_{-4}	s_4	s_4	s_{-2}	s_2	s_4
2	s_3	s_0	s_1	s_2	s_1	s_1	s_0	s_1	s_2
3	s_{-1}	s_2	s_2	s_3	s_2	s_1	s_0	s_2	s_3
4	s_2	s_2	s_1	s_1	s_3	s_4	s_1	s_2	s_1
5	s_4	s_4	s_{-4}	s_2	s_3	s_4	s_{-3}	s_2	s_4
6	s_1	s_3	s_2	s_{-1}	s_2	s_2	s_1	s_{-2}	s_2
7	s_3	s_4	s_3	s_4	s_3	s_3	s_3	s_{-4}	s_{-4}
8	s_4	s_4	s_{-3}	s_4	s_4	s_4	s_{-2}	s_{-3}	s_4

Table 4 Degrees of inclusion of condition concepts by decision concepts

j	Y_1	Y_2	Y_3
$\mathcal{D}(Y_k, C_j)$			
C_1	0.375	0.375	0.625
C_2	0.125	0.375	0.625
C_3	0.625	0.75	0.875
C_4	0.375	0.375	0.875
C_5	0.125	0.25	0.75
C_6	0.125	0.25	0.75

Table 5 Degrees of inclusion of decision concepts by condition concepts

j	Y_1	Y_2	Y_3
$\mathcal{D}(C_j, Y_k)$			
C_1	0.875	0.75	0.625
C_2	1	0.875	0.75
C_3	0.75	0.625	0.375
C_4	0.75	0.75	0.625
C_5	1	1	0.625
C_6	1	0.875	0.75

certainty as 1 in lower approximation of Y_2. Hence it can be seen that the lower approximation of Y_2 is C_3 with the degree of certainty as 0.75. Similarly in order to find the upper approximation of Y_2, we look for maximum degree of covering of Y_2 by any C_j in Table 5. We find that $\mathcal{D}(C_5, Y_2)$ is 1. In other words, the upper approximation of Y_2 is found to be C_5 with degree of certainty as 1.

That means that the linguistic coefficient of an alternative in the granule "not invest" should be at least as high as its respective coefficient against C_3 but not higher than its linguistic coefficient against C_5. The degree of certainty in this approximation is the minimum of the two degree of certainty values for the lower and upper approximation. In the present case the degree of certainty in this approximation is 0.75.

Approximation of Y_3

It is seen from Table 4 that the highest degree of covering of any C_j by Y_3 is 0.875 for C_3 and C_4. Therefore, the lower approximation of Y_3 is $C_3 \cap C_4$ with degree of certainty as 0.875. We do not find any C_j such that $\mathcal{D}\big(Y_3, C_3 \cap C_4 \cap C_j\big) = 1; j = 1, \ldots, 6; j \neq 3, 4.$ Hence it can be seen that the lower approximation of Y_3 is $C_3 \cap C_4$ with the degree of certainty as 0.875. Similarly, it can be seen from Table 5 that the maximum degree of covering of Y_3 by any C_j is 0.75 for C_2 and C_6. Also, it is found that $\mathcal{D}(C_2 \cup C_4 \cup C_6, Y_3) = 1$. Therefore, the upper approximation of Y_3 is $C_2 \cup C_4 \cup C_6$ with degree of certainty as 1. This infers that the linguistic assessment of an alternative in the granule "defer invest" is approximated to lie somewhere between those against C_3 and C_4, and maximum of the linguistic assessments against C_2, C_4 and C_6. The degree of certainty in this approximation is 0.875.

5.3 Comparison with fuzzy rough set

In fuzzy rough set (FRS), every sample x_i $(i \in [1, m])$ from a dataset of m objects is determined by N fuzzy characteristics. Each of these is a fuzzy set that is measured as partial membership degree, $\mu_i^j (j = 1, \ldots, N) \in [0, 1]$:

$$x_i = \left[\mu_i^1, \mu_i^2, \ldots, \mu_i^N\right]. \tag{27}$$

Hence, in the fuzzy rough set theory, each of the condition concepts is a fuzzy set. In order to apply the fuzzy rough set theory to our case-study at hand, we first generate the set of fuzzy condition concepts from the set of attributes C and the set of linguistic quantifiers S by obtaining their Cartesian product. As a result, the fuzzy condition concepts generated corresponding to C_1 are

$$\begin{aligned}&C_1(\text{Prospectivity}) = \{C_{11}: \text{extremely poor prospectivity},\\&C_{12}: \text{very poor prospectivity}, C_{13}: \text{poor prospectivity},\\&C_{14}: \text{slightly poor prospectivity}, C_{15}: \text{medium prospectivity},\\&C_{16}: \text{slightly good prospectivity}, C_{17}: \text{good prospectivity},\\&C_{18}: \text{very good prospectivity}, C_{19}: \text{extremely good prospectivity}\}\end{aligned}$$

Similarly, fuzzy condition concepts corresponding to remaining attributes $\{C_j\}, j = 2, \ldots, 6$ can also be generated. Hence, the number of condition concepts in an example like the case-study at hand is $N = |S| * |C|$, where $|.|$ refers to the cardinality of the set. In the present case-study, we need to deal with a set of $N = 54$ condition concepts.

As the next step, depending upon the expert's evaluation of each object x_i against a condition concept $C_{jk} (j = 1, \ldots, 6), (k = 1, \ldots, 9)$, a partial membership degree μ_i^{jk} is assigned. Thus, object x_i is characterized by the following membership degrees:

$$x_i = \left[\left(\mu_i^{11}, \ldots, \mu_i^{19}\right), \left(\mu_i^{21}, \ldots, \mu_i^{29}\right), \ldots, \left(\mu_i^{61}, \ldots, \mu_i^{69}\right)\right].$$

Once, we have the partial membership degrees $\left\{\mu_i^{jk}\right\} (j = 1, \ldots, 6)(k = 1, \ldots, 9)$ for the objects $\{x_i\}, i = 2, \ldots, m$, we could apply the standard fuzzy rough set framework as given in (7) and (8).

It can be easily observed that as the number of features increase in a dataset, the number of fuzzy condition concepts grow tremendously. Hence, the fuzzy rough set approach may be computationally too much demanding, when the number of attributes is large. Secondly, since approximations in FRS are based on the partial membership degrees of the given set of objects (that may be termed as training samples) in various condition concepts, there is no guarantee that these approximations hold good in general, for a larger population. Also, as we briefly mentioned earlier, the impact of even few noisy samples (with incorrect membership degree) is quite strong on the approximations obtained, for FRS is solely based on membership degrees.

In contrast, the proposed linguistic rough set (LRS) directly operates on linguistic quantifiers (which are fuzzy sets in themselves). As a result, the number of condition concepts is far lesser. In the case-study at hand, the information system is comprised of condition concepts with linguistic quantifiers as values. The approximations obtained are also in terms of these linguistic quantifiers. Since, LRS does not require numerical membership degrees; it is not having the difficulties associated with assigning membership grades to condition concepts and their interpretation. The linguistic quantifiers and the principle of inclusion degree at the heart of LRS ensure robust and generalized approximations. Unlike FRS, the trade-off between generalization and accuracy (to the given state of knowledge) in approximations obtained through LRS can be controlled by the adjustable inclusion degree.

6 Conclusions

Pointing to the complementarity of linguistic quantifiers and rough set theory, specifically against their common background of granulation of values drawn together by similarity and vagueness, we have proposed linguistic rough set (LRS). LRS provides a method to model the imprecision in natural languages while retaining the original imprecision. LRS is inspired by the human thought process that constantly makes use of granulation of values for information summarization in decision making with partial knowledge. The proposed LRS framework gives general and imprecise approximations of natural representations. The simplicity of operating directly with linguistic quantifiers adds to the usefulness of LRS in the real world decision making.

The measures for measuring the accuracy, generalization and crispness in the approximations obtained through LRS are proposed. A comprehensive approach is developed LRS to model natural representations. This approach is illustrated through a real world case-study. The proposed framework, despite its simplicity, is comprehensive enough to represent imprecise concepts and yield robust approximations. The work carried out in this study has several possible extensions, since this is the introductory paper on LRS. LRS can be integrated with probability theory to model occurrence related uncertainty along with uncertainties of imprecision and vagueness. In this regard, interesting extensions such as LRS based probabilistic classification, probabilistic LRS, and Bayesian LRS models can be developed. Also, on the lines of several extensions of rough set theory, it should be interesting to have decision-theoretic LRS, variable precision LRS, and parameterized LRS models. These possible extensions will foster the use of rough set theory in real world decision making and would broaden its domain of applications.

References

1. Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning, Part I. Inf Sci 8:199–249
2. Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning, Part II. Inf Sci 8:301–357
3. Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning, Part III. Inf Sci 9:43–80
4. Zadeh LA (1999) From computing with numbers to computing with words from manipulation of measurements to manipulation of perceptions. IEEE Trans Circuits Syst 45(1):105–119
5. Zadeh LA (2005) Toward a generalized theory of uncertainty (GTU)—an outline. Inf Sci 172(1–2):1–40
6. Zadeh LA (2006) Generalized theory of uncertainty (GTU)—principal concepts and ideas. Comput Stat Data Anal 51(1):15–46
7. Zadeh LA (2002) Toward a perception-based theory of probabilistic reasoning with imprecise probabilities. J Statistical Plan Inference 105(1):233–264
8. Zadeh LA (2000) Outline of a computational theory of perceptions based on computing with words. In: Sinha NK, Gupta MM (eds) Soft computing and intelligent systems: theory and applications. Academic Press, London, pp 3–22
9. Zadeh LA (2004) A note on web intelligence, world knowledge and fuzzy logic. Data Knowl Eng 50(3):291–304
10. Zadeh LA (2001) A new direction in AI: toward a computational theory of perceptions. AI Mag 22(1):73–84
11. Zadeh LA (1983) A computational approach to fuzzy quantifiers in natural languages. Comput Math Appl 9(1):149–184
12. Zadeh LA (1997) Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets Syst 90(2):111–127
13. Yager RR (1991) On linguistic summaries of data. In: Piatetsky-Shapiro G, Frawley B (eds) Knowledge discovery in databases. MIT Press, Cambridge, pp 347–363
14. Yager RR (1982) A new approach to the summarization of data. Inf Sci 28:69–86
15. Pawlak Z (1982) Rough sets. Int J Comput Inf Sci 11:341–356
16. Dubois D, Prade H (1990) Rough fuzzy sets and fuzzy rough sets. Int J Gen Syst 17:191–209
17. Radzikowska AM, Kerre EE (2002) A comparative study of fuzzy rough sets. Fuzzy Sets Syst 126(2):137–155
18. Jensen R, Shen Q (2004) Semantics-preserving dimensionality reduction: rough and fuzzy-rough based approaches. IEEE Trans Knowl Data Eng 16(12):1457–1471
19. Beauboef T, Petry FE, Arora G (1998) Information-theoretic measures of uncertainty for rough sets and rough relational databases. Inf Sci 109:185–195
20. Bodjanova S (1997) Approximation of fuzzy concepts in decision making. Fuzzy Sets Syst 85(1):23–29
21. Chunsheng F, Fending C, Qiang W, Shouchao W, Ju W (2010) Application of rough sets theory to sources analysis of atmospheric particulates, bioinformatics and biomedical engineering (iCBBE). In: 2010 4th International Conference, 18–20 June, pp 1–5

22. Shi C (2009) The application of rough sets and fuzzy sets in the supply chain knowledge sharing risk early-warning. Fuzzy systems and knowledge discovery. FSKD '09. In: Sixth International Conference, vol 6, 14–16 Aug, pp 441–445
23. Bordogna G, Fedrizzi M, Passi G (1997) A linguistic modeling of consensus in group decision making based on OWA operator. IEEE Trans Syst Man Cybernetics Part A Syst Hum 27:126–132
24. Levrat E, Voisin A, Bombardier S, Bremont J (1997) Subjective evaluation of car seat comfort with fuzzy set techniques. Int J Intell Syst 12:891–913
25. Zeshui Xu (2007) A method for multiple attribute decision making with incomplete weight information in linguistic setting. Knowl Based Syst 20(8):719–725
26. Zhang WX, Leung Y (1996) Theory of inclusion degrees and its applications to uncertainty inferences. Soft computing in intelligent systems and information processing, IEEE, New York, pp 496–501
27. Xu ZB, Liang JY, Dang CY, Chin KS (2002) Inclusion degree: a perspective on measures for rough set data analysis. Inf Sci 141:227–236
28. Liu Dun, Li Tianrui, Ruan Da (2011) Probabilistic model criteria with decision-theoretic rough sets. Inf Sci 181(17):3709–3722
29. Dombi J (1982) Basic concepts for a theory of evaluation: the aggregative operator. Eur J Oper Res 10(3):282–293
30. Lawry J, Tang Y (2010) Granular knowledge representation and inference using labels and label expressions. Fuzzy Syst IEEE Trans 18(3):500–514
31. Zadeh LA (1996) Fuzzy logic = computing with words. IEEE Trans Fuzzy Syst 4:103–111
32. Dubois D, Prade H (1987) Rough fuzzy sets and fuzzy rough sets. Fuzzy Sets Syst 23:3–18
33. Wang X, Tsang ECC, Zhao S, Chen D, Yeung DS (2007) Learning fuzzy rules from fuzzy samples based on rough set techniques. Inf Sci 177:4493–4514
34. Wang Xizhao, Zhai Junhai, Lu Shuxia (2008) Induction of multiple fuzzy decision trees based on rough set technique. Inf Sci 178(16):3188–3202
35. Lin TY (1997) Granular computing. Announcement of the BISC Special Interest Group on Granular Computing
36. Qian YH, Liang JY, Yao YY, Dang CY (2010) Incomplete mutigranulation rough set. IEEE Trans Syst Man Cybern Part A 20:420–430
37. Liu Xin, Qian Yuhua, Liang Jiye (2014) A rule-extraction framework under multigranulation rough sets. Int J Mach Learn Cybernet 5(2):319–326
38. Lin Guoping, Liang Jiye, Qian Yuhua (2014) Topological approach to multigranulation rough sets. Int J Mach Learn Cybernet 5(2):233–243
39. Fang Lian-Hua, Li Ke-Dian, Li Jin-Jin (2013) Properties of two types of covering-based rough sets. Int J Mach Learn Cybernet 4(6):685–691
40. Kang Xiangping, Li Deyu (2013) Dependency space, closure system and rough set theory. Int J Mach Learn Cybernet 4(6):595–599
41. Pawlak Z (1991) Rough sets. Theoretical aspects of reasoning about data, system theory, knowledge engineering and problem solving, vol 9, Kluwer, Dordrecht