Dear,

We are very grateful to you for your valuable comments and suggestions. We have carefully revised the paper in accordance with these comments and suggestions. The added and modified parts are shown in red in the revised manuscript (and changes are marked). The main revisions are as follows.

**Comment 1:** However the comparison of the proposed approach is only based on numerical results with existing benchmarks. It will be more rigorous if the authors are able to provide a theoretical comparison with similar World.

Response: Thank you very much for your valuable suggestion.

We have added a theoretical comparison in the subsection 3.3 as follows:

**On Pages 11 and 12:**

### 3.3. Comparison analysis with two representative reducts

It is known that the definition of reducts from the algebra view is usually equivalent to its definition from the information view in a general information system. What’s more, the relative reduct of a decision system in the information view includes that in the algebra view. Thus, Wang et al. [46] declared that any relative reduct of a decision system in the information view must be its relative reduct in the algebra view, so that some heuristic algorithms can be designed further using this conclusion. Based on the ideas of the classification in [46], the definition of reducts based on neighborhood roughness joint entropy of a neighborhood decision system should be developed from the algebra view and the information view in neighborhood rough set theory. For convenience, the reduct in Definition 13 is named as the neighborhood entropy reduct. Liu et al. [38] presented a reduct based on positive region in the neighborhood decision system similar with classical rough set model. This representative relative reduct based on positive region is called the algebra view of the neighborhood rough set theory. Chen et al. [20] defined information quantity similar to information entropy to evaluate the neighborhood classes, used the joint entropy gain to evaluate the significance of a selecting attribute, and proposed a representative joint entropy gain-based reduction algorithm, which is called a reduct in the information view of neighborhood rough sets.

Given a neighborhood decision system \( NDS = \langle U, C \cup D, \delta \rangle \) with non-empty infinite set \( U \), \( B \subseteq C \) and \( D = \{d\} \). Then, a positive region reduct of the neighborhood decision system is presented as follows in [38]: for any \( a \in B \), if \( |POS(D)| = |POS_C(D)| \) and \( |POS_{B \setminus \{a\}}(D)| < |POS(D)| \), where \( POS(D) = \bigcup\{B(X) \mid X \in U / D\} \) is the positive region of \( D \) with respect to \( B \), \( B \) is a relative reduct of the neighborhood decision system.

**Proposition 7.** Given a neighborhood decision system \( NDS = \langle U, C \cup D, \delta \rangle \) with non-empty infinite set \( U \), and \( B \subseteq C \), if \( B \) is a neighborhood entropy reduct of the neighborhood decision system, then \( B \) is a positive region reduct of the neighborhood decision system.

**Proof.** Let \( U = \{x_1, x_2, \ldots, x_n, \ldots\} \), and \( U / D = \{d_1, d_2, \ldots, d_n \ldots\} \). Suppose that for a subset \( B \subseteq C \), it follows from Definition 13 that if \( NRH(D, B) = NRH(D, C) \), and for any \( a \in B \), there exists \( NRH(D, B) > NRH(D, B \setminus \{a\}) \), then \( B \) is a neighborhood entropy reduct of \( C \) relative to \( D \).
When \( \text{NRH}(D, B) = \text{NRH}(D, C) \), it can be obtained from Proposition 6 that \( n^a_x(x) = n^c_x(x), \)
\( \gamma^a_x(d_j) = \gamma^c_x(d_j) \) and \( n^a_y(x) \cap d_j = n^c_y(x) \cap d_j \) hold, where any \( x \in U \) and \( 1 \leq j \). By Eq. (13), one has that \( B(D)_a = C(D)_c \). So it is obvious that \( \text{POS}_a(D) = \text{POS}_c(D) \), i.e., \( |\text{POS}_a(D)| = |\text{POS}_c(D)| \).
For any \( a \in B \), \( B - \{a\} \subset B \), and from Theorem 1 in [38], one has that \( \overline{B - \{a\}}(D)_a \leq B(D)_a \), so that \( \text{POS}_{\text{ad}}(D) \subset \text{POS}(D) \) holds. Because for any \( a \in B \), there exists \( \text{NRH}(D, B) > \text{NRH}(D, B - \{a\}) \), thus \( \overline{B - \{a\}}(D)_a \leq B(D)_a \) holds. It follows that \( \text{POS}_{\text{ad}}(D) \subset \text{POS}(D) \). Thus, \( |\text{POS}_{\text{ad}}(D)| < |\text{POS}(D)| \) for any \( a \in B \). Therefore, \( B \) is a positive region reduct of the neighborhood decision system.

Notably, the inverse relation of this proposition generally does not hold. According to the above discussions, Proposition 7 shows that the definition of the neighborhood entropy reduct includes that of positive region reduct in the algebra view.

Given a neighborhood decision system \( \text{NDS} = \langle U, C \cup D, \delta \rangle \) with non-empty infinite set \( U \), \( B \subseteq C \) and \( D = \{d\} \). For any \( a \in B \), a reduct of the neighborhood decision system, named as the entropy gain reduct is proposed in [20] as follows: if \( H(Bd) = H(Cd) \) and \( H([B - \{a\}]d) < H(Bd) \), the entropy gain reduct of the neighborhood decision system.

\[ \text{Proposition 8.} \quad \text{Given a neighborhood decision system } \text{NDS} = \langle U, C \cup D, \delta \rangle \text{ with non-empty infinite set } U \text{ and } B \subseteq C, \text{ then } B \text{ is a neighborhood entropy reduct of the neighborhood decision system if and only if } B \text{ is an entropy gain reduct of the neighborhood decision system.} \]

**Proof.** \( \Rightarrow \) Let \( U = \{x_1, x_2, \ldots, x_n, \ldots\} \), and \( U/D = \{d_1, d_2, \ldots, d_n, \ldots\} \). Suppose that for a subset \( B \subseteq C \), it follows from Definition 13 that if \( \text{NRH}(D, B) = \text{NRH}(D, C) \), and for any \( a \in B \), there exists \( \text{NRH}(D, B) > \text{NRH}(D, B - \{a\}) \), then \( B \) is a neighborhood entropy reduct of \( C \) relative to \( D \). Similar to the proof of Proposition 7, when \( \text{NRH}(D, B) = \text{NRH}(D, C) \), from Proposition 6, one has that \( n^a_y(x) \cap d_j = n^c_y(x) \cap d_j \), where any \( x \in U \) and \( 1 \leq j \). It is obvious that \( H(Bd) = H(Cd) \). Since \( B - \{a\} \subset B \), from Proposition 2 in [20], one has that \( H([B - \{a\}]d) < H(Bd) \). Because for any \( a \in B \), there exists \( \text{NRH}(D, B) > \text{NRH}(D, B - \{a\}) \), so \( H([B - \{a\}]d) < H(Bd) \) holds. Hence, \( B \) is an entropy gain reduct of the neighborhood decision system.

\( \Leftarrow \) Suppose that for a subset \( B \subseteq C \), and any \( a \in B \), if \( H(Bd) = H(Cd) \) and \( H([B - \{a\}]d) < H(Bd) \), then \( B \) is an entropy gain reduct of \( C \) relative to \( D \). Similar to the proof of Proposition 6, when \( n^a_y(x) = n^c_y(x) \), by Eqs. (16), (17) and (19), one has that \( \gamma^a_y(d_j) = \gamma^c_y(d_j) \), and then it is obvious that \( n^a_y(x) \cap d_j = n^c_y(x) \cap d_j \), where any \( x \in U \) and \( 1 \leq j \). Thus, it can be obtained from Eq. (24) that \( \text{NRH}(D, B) = \text{NRH}(D, C) \). Because \( B - \{a\} \subset B \), it follows from Proposition 6 that \( \text{NRH}(D, B - \{a\}) \leq \text{NRH}(D, B) \). Since for any \( a \in B \), there exists \( H([B - \{a\}]d) < H(Bd) \). So, one has that \( \text{NRH}(D, B - \{a\}) < \text{NRH}(D, B) \). Therefore, \( B \) is a neighborhood entropy reduct of the neighborhood decision system.

Proposition 8 shows that in a neighborhood decision system, the neighborhood entropy reduct is equivalent to the entropy gain reduct in the information view. According to Proposition 7 and 8, it can be concluded that the definition of neighborhood entropy reduct includes two representative reducts proposed in the algebra view and the information view. Therefore, the definition of neighborhood entropy reduct denotes a mathematical quantitative measure to evaluate the knowledge uncertainty of different attribute sets in neighborhood decision systems.
Comment 2: Particularly, it is required to clearly explain the improvement made on the information and algebraic viewpoints.
Response: Thank you very much for your valuable suggestion.

We have given this explanation of the improvement on the information and algebraic viewpoints as follows:

**On Page 3:** As we can see, many existing methods of attribute reduction in neighborhood rough sets usually only start from the algebraic point of view or the information point of view, while the definition of attribute significance based on algebraic view only describes the effect of attributes on the subset of classification contained [46].

Although these methods have their own advantages, they are still inefficient and not suitable for reducing large-scale high-dimensional data, and the enhanced algorithms only decrease the computation time to a certain extent [47]. Inspired by this, to study neighborhood rough sets from the two views and achieve great uncertainty measures in neighborhood decision systems, the algebra view and the information view will be combined to develop attribute reduction algorithm for infinite sets in continuous-valued data sets.

**On Page 10:** It is noted that Wang et al. [46] stated that all conceptions and computations in rough set theory based on the upper and lower approximation sets are called the algebra view of the rough set theory, and the notions of information entropy and its extensions are called the information view of rough sets. It follows from Eq. (24) that $r^a(d_i)$ is the neighborhood roughness of $d_i$ with respect to $B$ in the algebra view and it represents the degree of incompleteness of obtaining knowledge of set $d_i$, and

$$\mathop{\sum}_{j=1}^{m} \int_{x \in U} \log_{2} \left( \frac{m(B_i \cap d_j)}{m(U)} \right) dx$$

is the definition of joint entropy in the information view. Hence, Definition 8 can efficiently analyze and measure the uncertainty of neighborhood decision systems based on Lebesgue and entropy measures from both the algebra view and the information view.

Comment 3: The authors must improve the literature review for a more sound comparison of their approach using Lebesgue and entropy measures combining algebra view with information view. The following reference may be relevant:
Response: Thank you very much for your valuable suggestion.

We have improved the literature review and added the assigned relevant references as follows:

**First, on Page 2:** Halčinová et al. [28] investigated the weighted Lebesgue integral by Lebesgue differentiation theorem, and used Lebesgue measure to develop the standard weighted $L^p$-based sizes. Park et al. [29] expressed a measurable map through Lebesgue integration to define the Cumulative Distribution Transform of density functions. Recently, scholars [30-32] introduced Lebesgue measure as a promising additional method to resolve some problems in the application of data analysis.

**On Page 26:**


Second, to the best of my knowledge, Halcinová et al. [28] used Lebesgue measure to develop the standard weighted $L_p$-based sizes. Park et al. [29] expressed a measurable map through Lebesgue integration to define the Cumulative Distribution Transform of density functions. Marzio et al. [30] only employed the Lebesgue measure as parametrization of a point to denote the tangent-normal decomposition in the unit hypersphere. Shiraz et al. [31] introduced Lebesgue measure to only study the upper trust and the lower trust in rough space. Zhang et al. [32] used Borel measurable function as the hypothesis to guarantee the existence of Lebesgue–Stieltjes integral.

In our manuscripts, since the Lebesgue measure can efficiently evaluate infinite sets, it is therefore necessary to employ Lebesgue measure to study uncertainty measures and efficient reduction algorithms for infinite sets in neighborhood decision systems. Then, the Lebesgue measure [25] is introduced into neighborhood entropy to investigate the uncertainty measures in neighborhood decision systems, an attribute reduction method using Lebesgue and entropy measures is presented, and then a heuristic search algorithm is designed to analyze the uncertainty and noisy of continuous and complex data sets.

In general, since our motivation is different from those of [28-32], the comparison with the other approaches using Lebesgue measure in [28-32] is very difficult to be obtained.

Thank you once again for your constructive and valuable comments.

Best wishes,

Prof. Jiucheng Xu, Ph.D.,
On behalf of all authors
College of Computer and Information Engineering, Henan Normal University
Email: jiuchxu@gmail.com