Complex & Intelligent Systems (2022) 8:5309–5328
https://doi.org/10.1007/s40747-022-00763-0

ORIGINAL ARTICLE

Online group streaming feature selection using entropy-based uncertainty measures for fuzzy neighborhood rough sets

Jiucheng Xu[1,2] · Yuanhao Sun[1,2] · Kanglin Qu[1,2] · Xiangru Meng[1,2] · Qinchen Hou[1,2]

Received: 13 January 2022 / Accepted: 22 April 2022 / Published online: 16 May 2022

Abstract

Online group streaming feature selection, as an essential online processing method, can deal with dynamic feature selection tasks by considering the original group structure information of the features. Due to the fuzziness and uncertainty of the feature stream, some existing methods are unstable and yield low predictive accuracy. To address these issues, this paper presents a novel online group streaming feature selection method (FNE-OGSFS) using fuzzy neighborhood entropy-based uncertainty measures. First, a separability measure integrating the dependency degree with the coincidence degree is proposed and introduced into the fuzzy neighborhood rough sets model to define a new fuzzy neighborhood entropy. Second, inspired by both algebra and information views, some fuzzy neighborhood entropy-based uncertainty measures are investigated and some properties are derived. Furthermore, the optimal features in the group are selected to flow into the feature space according to the significance of features, and the features with interactions are left. Then, all selected features are re-evaluated by the Lasso model to discard the redundant features. Finally, an online group streaming feature selection algorithm is designed. Experimental results compared with eight representative methods on thirteen datasets show that FNE-OGSFS can achieve better comprehensive performance.

Keywords Streaming feature selection · Streaming groups · Fuzzy neighborhood rough sets · Coincidence degree · Fuzzy neighborhood entropy · Uncertainty measures

Introduction

Feature selection, as an important data preprocessing technique, plays a key role in knowledge discovery, pattern recognition, and machine learning [1–5]. It aims to select the optimal subset to improve classification accuracy and reduce computational complexity. Traditional feature selection methods are batch methods, assuming that the feature space is fixed [6]. However, the complete feature space is not available in real-world situations, such as high-resolution planetary image analysis during Mars rover operations [7], where obtaining the entire feature set in this scenario means using image data covering the entire surface of Mars, which is clearly not feasible. Therefore, dynamic feature selection has attracted the continuous attention of scholars in recent years [8–12].

✉ Yuanhao Sun
sunyuanhaohtu@163.com

Jiucheng Xu
jiuchxu@gmail.com

Kanglin Qu
klinqu@163.com

Xiangru Meng
13837169720@163.com

Qinchen Hou
15560122618@163.com

1 College of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, China

2 Engineering Lab of Intelligence Business and Internet of Things, Henan Normal University, Xinxiang 453007, Henan, China

Related work

Dynamic feature selection can be divided into feature selection with feature streams and data streams [13–18]. Online feature selection with a feature stream assumes that features flow into the feature space in batches over time, which can be further segmented into individual streaming feature selection and group streaming feature selection based on the structural information of the features [19–22].

Online individual streaming feature selection is characterized by features arriving in the feature space one by one [23–28]. Perkins et al. [29] first implemented feature selection in an online manner using a fast gradient-based heuristic. Wu et al. [30] proposed a streaming feature selection framework consisting of both online correlation analysis and online redundancy analysis. Zhou et al. [31] considered the interactions between features and efficiently selected features with interactions. Eskandari et al. [32] first proposed a new streaming feature selection method based on rough sets theory and improved it in [33]. Lin et al. [34] introduced fuzzy mutual information in multilabel learning to evaluate the quality of features for streaming feature selection scenarios. You et al. [35] proposed a causal feature selection algorithm for online streaming feature selection scenarios. The above methods can only handle the scenarios where features arrive one by one but ignore the original group structure of the features [36]. For example, in the fields of drug localization [37] and image analysis [38], where features mostly arrive in the feature space as groups.

Online group streaming feature selection can consider the structural information of the features, and thus achieve better classification results [39,40]. Li et al. [40] used entropy and mutual information to perform online group streaming feature selection. Wang et al. [41] proposed a framework for online feature selection using prior knowledge of group structure information, including intra-group feature selection and inter-group feature selection. Yu et al. [42] extended the scalable and accurate online feature selection approach to handle the group feature selection problem in the sparse case. Liu et al. [43] proposed a new framework based on group structure analysis for online multilabel group streaming feature selection. Zhou et al. [44] designed a new online group streaming feature selection algorithm focused on feature interaction based on mutual information. Unfortunately, the information present in the real world contains many subjective concepts, such as pretty, young, and moral, these concepts have no clear boundaries and thus create ambiguity and uncertainty [45–50]. Existing streaming feature selection methods cannot deal well with tasks in the context of fuzziness and uncertainty.

The classification task learns a classification model, i.e., a classifier, from the existing training samples [51–53]. When test data arrive, predictions can be made based on the classifier to map the new data items to one of the classes in the given category [52]. Recently, feature selection based on rough sets and fuzzy sets from algebra and information views has been frequently reported to measure uncertainty in classification tasks [54–56]. Wang et al. [57] proposed a fuzzy neighborhood rough sets model (FNRS) using parameterized fuzzy neighborhood information granules that can effectively prevent the effect of noise. Shreevastava et al. [58] proposed a new intuitionistic fuzzy neighborhood rough sets model for heterogeneous datasets, which combined intuitionistic fuzzy sets and neighborhood rough sets. An et al. [53] proposed a relative fuzzy rough set model and designed a classifier based on the maximum positive domain for the problem of large differences in the class density of the data distribution. The above studies discussed feature selection from an algebraic view, where the significance of features can only state the consequence of features contained in the feature subset [55,59]. Sang et al. [60] proposed a fuzzy dominant neighborhood rough sets model for possible noisy data in biased-ordered information systems. Xu et al. [61] redefined fuzzy neighborhood relations and introduced them into conditional entropy, proposing a new fuzzy neighborhood conditional entropy. Zhang et al. [62] proposed active incremental feature selection using the information entropy of introduced instances based on fuzzy rough sets. Nevertheless, the feature significance of these information view-based references merely interprets the influence of uncertainty classification on features [54,62]. Sun et al. [63] combined the fuzzy neighborhood rough sets with the neighborhood multigranulation rough sets and proposed a fuzzy neighborhood multigranulation rough sets model. Xu et al. [64] fused the self-information measure into the fuzzy neighborhood in the upper and lower approximations and proposed a fuzzy neighborhood joint entropy based on fuzzy neighborhood self-information. Xu et al. [65] defined multilabel fuzzy neighborhood conditional entropy and approximation accuracy to solve the classification problem under a multilabel decision system. It is confirmed that the combination of algebra and information views can make the measurement mechanism more comprehensive [63,65].

Fuzzy rough sets theory and its applications are also widely used to solve some practical problems [66–72]. Xu et al. [66] proposed a fuzzy rough uncertainty measure model for tumor diagnosis and microarray gene selection. Liu et al. [68] proposed a co-evolutionary model for crime inference based on fuzzy rough sets. These reported studies evaluate features by the consistency of conditional and decision features in information granularity, ignoring the separability of decision information granularity for different conditional features. Nonetheless, the separability of conditional features is closely related to their performance in classification tasks [73,74]. Hu et al. [73] defined the aggregation degree of intraclass objects and the dispersion degree of between-class objects to measure the significance of features. The separability-based evaluation function has a profound impact on the improvement of accuracy and time efficiency. However, these fuzzy rough sets-based approaches are only applicable to traditional feature selection and cannot deal with feature selection in dynamic environments.

Our work

Motivated by this, to effectively deal with the streaming feature selection task in fuzzy and uncertainty contexts, this paper proposes some fuzzy neighborhood entropy-based uncertainty measures and investigates a novel online group streaming feature selection method, named FNE-OGSFS. The main innovation points are as follows:

- To better evaluate the classification quality of features in terms of separability, we define a new separability degree (SD) by integrating the coincidence degree and the dependency degree for fuzzy neighborhood rough sets and fuse it with FNRS to define a new fuzzy neighborhood entropy. Then, we propose the concepts of fuzzy neighborhood joint entropy, fuzzy neighborhood conditional entropy and fuzzy neighborhood mutual information. The related properties are explored and proven.
- To better discuss the measure of online streaming feature selection from both algebra and information views, we propose fuzzy neighborhood symmetric uncertainty. Then, we present a series of uncertainty measures such as the significance, fuzzy neighborhood interaction gain and contrast ratio. The related theorems are derived and proven. Furthermore, we construct an online group streaming decision system to retain features with strong approximation ability when features dynamically flow into the feature space while removing redundant features.
- Based on this, we design a new online group streaming feature selection algorithm, named FNE-OGSFS. First, the significance is used for intra-group feature selection. Second, online interaction analysis is performed on feature groups flowing into the feature space based on the fuzzy neighborhood interaction gain and contrast ratio. Finally, redundant features are removed using the Lasso model. Experimental results on thirteen different types of real-world datasets confirmed that FNE-OGSFS can effectively select the optimal feature subset.

The remainder of this paper is organized as follows. "Preliminaries" reviews the related knowledge of FNRS and the coincidence degree. "Fuzzy neighborhood entropy-based uncertainty measures" presents the separability degree and some fuzzy neighborhood entropy-based uncertainty measures. "Online group streaming feature selection approach" develops a novel online group streaming feature selection approach. "Experimental results" provides the experimental analysis on thirteen datasets. "Conclusion and future work" concludes the paper with an outlook on the future.

Preliminaries

The FNRS is an effective model for feature selection and knowledge discovery. In this section, we review some basic concepts of fuzzy neighborhood rough sets. In addition, we introduce some basic knowledge related to the coincidence degree to facilitate the subsequent discussions.

Fuzzy neighborhood rough sets

Let $DS = \langle U, C, D \rangle$ be a decision system, where $U = \{x_1, x_2, \ldots, x_n\}$ is a nonempty finite set called the theoretical domain, C is the set of conditional features of the sample, and D is the set of decision features of the sample. $U/D = \{D_1, D_2, \cdots, D_l\}$ means D divides U into l equivalence classes.

Let $A \subseteq C$ be a subset of conditional features on U, and a fuzzy binary relation R_A can be induced by A. Then, R_A is called a fuzzy similarity relation when it satisfies both reflexivity and symmetry [57]:

(1) Reflexivity: $R_A(x, x) = 1, \forall x \in U$.
(2) Symmetry: $R_A(x, y) = R_A(y, x), \forall x, y \in U$.

Let $a \in A$ and R_a be the fuzzy similarity relation obtained by a. Then, R_A can be expressed as $R_A = \bigcap_{\forall a \in A} R_a$.

Definition 1 Given $DS = \langle U, C, D \rangle$, let the fuzzy neighborhood radius parameter be δ $(0 < \delta \leq 1)$, which is used to describe the similarity of samples, and for any $x, y \in U, a \in C$, the fuzzy neighborhood similarity relation between samples x and y with respect to feature a is denoted as

$$R_a(x, y) = \begin{cases} 0, & \Delta > \delta \\ 1 - \Delta, & \Delta \leq \delta \end{cases}, \tag{1}$$

where Δ denotes distance and equals to $|f(a, x) - f(a, y)|$.

Definition 2 Given $DS = \langle U, C, D \rangle$, the fuzzy neighborhood similarity matrix of samples x and y with respect to feature a is $[x]_a^\delta(y) = R_a(x, y)$, and for any $A \subseteq C$, $[x]_A^\delta(y) = \min_{\forall a \in A}([x]_a^\delta(y))$, then the parameterized fuzzy neighborhood information granule of x with respect to A is expressed as

$$\delta_A(x) = \begin{cases} 0, & [x]_A^\delta(y) < 1 - \delta \\ [x]_A^\delta(y), & [x]_A^\delta(y) \geq 1 - \delta \end{cases} \tag{2}$$

Definition 3 Given $DS = \langle U, C, D \rangle$, $A \subseteq C$, and $U/D = \{D_1, D_2, \cdots, D_l\}$, the fuzzy decision derived from D is denoted as $FD = \left\{FD_1^T, FD_2^T, \cdots, FD_l^T\right\}$, where $FD_j = \left\{FD_j(x_1), FD_j(x_2), \ldots, FD_j(x_n)\right\}$ is the fuzzy equivalence class of sample decisions. $FD_j(x)$ is the degree of membership and denoted by

$$FD_j(x) = \frac{|\delta_A(x) \cap D_j|}{|\delta_A(x)|}, \quad j = 1, 2, \ldots, l, \tag{3}$$

where $|\cdot|$ represents the cardinality.

Definition 4 Let A, B be two fuzzy sets on U. Then, the inclusion degree of A on B is expressed as

$$\text{Inc}(A, B) = \frac{|A \subseteq B|}{|U|}, \tag{4}$$

where $|A \subseteq B|$ denotes the number of samples whose membership degree on A is not greater than that on B.

Definition 5 Given $DS = \langle U, C, D\rangle$, let β be a variable precision parameter, $A \subseteq C$ and $X \subseteq U$. Then, the fuzzy neighborhood upper and lower approximations of X with respect to A are denoted, respectively, by

$$\overline{FN_A^{\beta}}(X) = \{x \in U | \text{Inc}(\delta_A(x), X) \geq \beta\}, \tag{5}$$

$$\underline{FN_A^{\delta}}(X) = \{x \in U | \text{Inc}(\delta_A(x), X) \geq \delta\}. \tag{6}$$

Definition 6 Given $DS = \langle U, C, D\rangle$, $A \subseteq C$, the fuzzy decision generated by D is $FD = \{FD_1^T, FD_2^T, \cdots, FD_l^T\}$. Then, the fuzzy neighborhood positive region of D with respect to A is denoted as

$$\text{POS}_A^{\delta}(D) = \bigcup_{j=1}^{l} \underline{FN_A^{\delta}}(FD_j). \tag{7}$$

Definition 7 Given $DS = \langle U, C, D\rangle$, $A \subseteq C$, the fuzzy neighborhood dependency degree of D in relation to A is expressed as

$$\text{Dep}_A^{\delta}(D) = \frac{|\text{POS}_A^{\delta}(D)|}{|U|}. \tag{8}$$

Coincidence degree

The purpose of feature selection is to retain the features with high separability and strong approximation ability and to remove the trivial features [73]. If the coincidence degree of the original data from different categories is high and the coincidence degree of the selected data is low, then the importance of the retained features is high.

Definition 8 Given $DS = \langle U, C, D\rangle$, $A \subseteq C$, $a \in C$, $U/D = \{D_1, D_2, \ldots, D_l\}$, $D_i, D_j \in U/D$, then the coincidence degree of a in regard to D_i and D_j is defined as

$$\text{Coin}(a|D_i, D_j) = \frac{\left|\left[m_{D_i}^a, M_{D_i}^a\right] \cap \left[m_{D_j}^a, M_{D_j}^a\right]\right| + \vartheta}{\left|\left[m_{D_i}^a, M_{D_i}^a\right] \cup \left[m_{D_j}^a, M_{D_j}^a\right]\right| + \vartheta}, \tag{9}$$

where $m_{D_i}^a = \min_{x \in D_i} f(x, a)$,and $M_{D_i}^a = \max_{x \in D_i} f(x, a)$.

When $\left|\left[m_{D_i}^a, M_{D_i}^a\right] \cup \left[m_{D_j}^a, M_{D_j}^a\right]\right| = 0$ or $\left|\left[m_{D_i}^a, M_{D_i}^a\right] \cap \left[m_{D_j}^a, M_{D_j}^a\right]\right| = 0$, ϑ is a very small positive constant, and in other cases, $\vartheta = 0$.

Definition 9 Let d be the number of $D_i \neq D_j$; then, the coincidence degree of D with respect to a is expressed as

$$\text{CD}(a) = \frac{1}{d} \sum_{D_i \neq D_j} \text{Coin}(a|D_i, D_j). \tag{10}$$

$\text{CD}(a)$ represents the coincidence degree of samples between different categories and evaluates the classification ability of feature a in terms of separability. In the process of feature selection, we need to select the feature that can decrease the coincidence degree.

Fuzzy neighborhood entropy-based uncertainty measures

To select features with high separability and strong approximation ability, this section defines a new separability measure and a new fuzzy neighborhood entropy. The feature selection method combining algebraic view and information view can achieve better classification results, and from this perspective, an uncertainty measure based on fuzzy neighborhood entropy is constructed and some related properties are derived.

Definition 10 Given $DS = \langle U, C, D\rangle$, $a \in C$, the separability degree of D with respect to a is expressed as

$$SD_a^{\delta}(D) = \left(1 - \text{Dep}_a^{\delta}(D)\right) * CD(a). \tag{11}$$

Let $A \subseteq C$ and $A = \{A_1, A_2, \ldots, A_r\}$; then, the separability degree of D in regard to A is denoted by

$$SD_A^{\delta}(D) = \frac{1}{r} \sum_{i=1}^{r} SD_{A_i}^{\delta}(D). \tag{12}$$

Property 1 *The smaller the value of $SD_A^{\delta}(D)$ is, the more important feature A is.*

Proof Obviously, $SD_A^\delta(D)$ is only related to two parts: $\mathrm{Dep}_A^\delta(D)$ and $\mathrm{CD}(A)$. From the properties of FNRS and Definition 10, it can be obtained that the larger the value of $\mathrm{Dep}_A^\delta(D)$ is, the more important feature A is, and the smaller the value of $SD_A^\delta(D)$ is. Moreover, the nature of the coincidence degree shows that we need to choose the features that can reduce the overlap. In brief, the smaller the value of $SD_A^\delta(D)$ is, the higher the separability of A and the more important feature A is.

Property 2 $0 < SD_A^\delta(D) \le 1$.

Proof It follows from Definition 7 that $0 \le Dep_a^\delta(D) \le 1$; thus, we can see that $0 \le \left(1 - Dep_a^\delta(D)\right) \le 1$. By Definition 8, we have $0 < Coin\left(a|D_i, D_j\right) \le 1$; thus, $0 < CD(a) \le 1$. Hence, we have $0 < SD_a^\delta(D) \le 1$ from Definition 10; furthermore, $0 < SD_A^\delta(D) \le 1$.

Definition 11 Given $DS = \langle U, C, D\rangle$, $U = \{x_1, x_2, \ldots, x_n\}$, and $A \subseteq C$, the fuzzy neighborhood entropy of A is defined as

$$\mathrm{FNE}_\delta(A) = -\frac{1}{|U|}\sum_{k=1}^{n}\log_2\frac{|\delta_A(x_k)| * SD_A^\delta(D)}{|U|}. \tag{13}$$

Definition 12 Given $DS = \langle U, C, D\rangle$, $A, B \subseteq C$, the fuzzy neighborhood joint entropy of A and B is denoted by

$$\mathrm{FNE}_\delta(A, B) = -\frac{1}{|U|} \times \sum_{k=1}^{n}\frac{|\delta_A(x_k) \cap \delta_B(x_k)| * SD_{A\cup B}^\delta(D)}{|U|}. \tag{14}$$

Definition 13 Given $DS = \langle U, C, D\rangle$, $A, B \subseteq C$, the fuzzy neighborhood conditional entropy of A with respect to B is denoted as

$$\mathrm{FNE}_\delta(A|B) = -\frac{1}{|U|} \times \sum_{k=1}^{n}\log_2\frac{|\delta_A(x_k) \cap \delta_B(x_k)| * SD_{A\cup B}^\delta(D)}{|\delta_B(x_k)| * SD_B^\delta(D)}. \tag{15}$$

Definition 14 Given $DS = \langle U, C, D\rangle$, $A, B \subseteq C$, the fuzzy neighborhood mutual information of A and B is represented as

$$\mathrm{FNMI}_\delta(A; B) = -\frac{1}{|U|} \times \sum_{k=1}^{n}\log_2\frac{|\delta_A(x_k)|\,|\delta_B(x_k)| * SD_A^\delta(D) * SD_B^\delta(D)}{|U| * |\delta_A(x_k) \cap \delta_B(x_k)| * SD_{A\cup B}^\delta(D)}. \tag{16}$$

Property 3 *(1)* $FNMI_\delta(A; B) = FNMI_\delta(B; A)$.

(2) $FNMI_\delta(A; B) = FNE_\delta(A) + FNE_\delta(B) - FNE_\delta(A, B)$.

(3) $FNMI_\delta(A; B) = FNE_\delta(A) - FNE_\delta(A|B) = FNE_\delta(B) - FNE_\delta(B|A)$.

Proof (1)

$$\begin{aligned}\mathrm{FNMI}_\delta(A; B) &= -\frac{1}{|U|}\sum_{k=1}^{n}\log_2\frac{|\delta_A(x_k)|\,|\delta_B(x_k)| * SD_A^\delta(D) * SD_B^\delta(D)}{|U| * |\delta_A(x_k) \cap \delta_B(x_k)| * SD_{A\cup B}^\delta(D)}\\ &= -\frac{1}{|U|}\sum_{k=1}^{n}\log_2\frac{|\delta_B(x_k)|\,|\delta_A(x_k)| * SD_B^\delta(D) * SD_A^\delta(D)}{|U| * |\delta_B(x_k) \cap \delta_A(x_k)| * SD_{B\cup A}^\delta(D)}\\ &= \mathrm{FNMI}_\delta(B; A).\end{aligned}$$

(2)

$$\begin{aligned}FNMI_\delta(A; B) &= -\frac{1}{|U|}\sum_{k=1}^{n}\log_2\frac{|\delta_A(x_k)|\,|\delta_B(x_k)| * SD_A^\delta(D) * SD_B^\delta(D)}{|U| * |\delta_A(x_k) \cap \delta_B(x_k)| * SD_{A\cup B}^\delta(D)}\\ &= \left(-\frac{1}{|U|}\sum_{k=1}^{n}\log_2\frac{|\delta_A(x_k)| * SD_A^\delta(D)}{|U|}\right)\\ &\quad + \left(-\frac{1}{|U|}\sum_{k=1}^{n}\log_2\frac{|\delta_B(x_k)| * SD_B^\delta(D)}{|U|}\right)\\ &\quad - \left(-\frac{1}{|U|}\sum_{k=1}^{n}\log_2\frac{|\delta_A(x_k) \cap \delta_B(x_k)| * SD_{A\cup B}^\delta(D)}{|U|}\right)\\ &= \mathrm{FNE}_\delta(A) + \mathrm{FNE}_\delta(B) - \mathrm{FNE}_\delta(A, B).\end{aligned}$$

(3)

$$\begin{aligned}\mathrm{FNMI}_\delta(A; B) &= -\frac{1}{|U|}\sum_{k=1}^{n}\log_2\frac{|\delta_A(x_k)|\,|\delta_B(x_k)| * SD_A^\delta(D) * SD_B^\delta(D)}{|U| * |\delta_A(x_k) \cap \delta_B(x_k)| * SD_{A\cup B}^\delta(D)}\\ &= \left(-\frac{1}{|U|}\sum_{k=1}^{n}\log_2\frac{|\delta_A(x_k)| * SD_A^\delta(D)}{|U|}\right)\\ &\quad - \left(-\frac{1}{|U|}\sum_{k=1}^{n}\log_2\frac{|\delta_A(x_k) \cap \delta_B(x_k)| * SD_{A\cup B}^\delta(D)}{|\delta_B(x_k)| * SD_B^\delta(D)}\right)\\ &= \mathrm{FNE}_\delta(A) - \mathrm{FNE}_\delta(A|B).\end{aligned}$$

Similarly, $FNMI_\delta(A; B) = FNE_\delta(B) - FNE_\delta(B|A)$.

Definition 15 Given $DS = \langle U, C, D\rangle$, $A, B \subseteq C$, the fuzzy neighborhood symmetrical uncertainty of A and B is represented as

$$\mathrm{FNSU}_\delta(A; B) = \frac{2\mathrm{FNMI}_\delta(A; B)}{FNE(A) + FNE(B)}. \tag{17}$$

In particular, let $FNSU_\delta(A; D) = FNMI_\delta(A; D)/FNE(A, D)$ when $B = D$.

Remark 1 Fuzzy neighborhood symmetrical uncertainty as a measure of uncertainty can better measure the significance of features. From Definition 15, we can see that $SD_A^\delta(D)$ denotes the separability degree from an algebraic view, and $\mathrm{FNMI}_\delta(A;B)$ represents the fuzzy neighborhood mutual information from an information view. Hence, $\mathrm{FNSU}_\delta(A;B)$ can measure the uncertainty from both algebra and information views.

Online group streaming feature selection approach

In this section, we first give a formalization of the problem of online group flow feature selection. Then, a new online group streaming feature selection algorithm is proposed based on the uncertainty measure proposed in the previous section, which includes three parts: intra-group feature selection, online interaction analysis, and online redundancy analysis. Finally, we perform a time complexity analysis of the proposed algorithm.

Problem formalization

Let $\mathrm{OGDS} = (U, G, D, h, t)$ be an online group streaming decision system, $U = \{x_1, x_2, \ldots, x_n\}$ be the set of samples, $G = \{G_1, G_2, .., G_w\}$ is the set of stream features, $G_i = \{f_1, f_2, \ldots, f_{m_t}\}^T$ is a set of features in G containing m_t features, and a new set of features G_t is obtained with unknown feature space at each stamp t. D is the set of decision features, h is the feature-to-class mapping function, and t is the time stamp. The problem of online group streaming feature selection is to select an optimal feature subset S in the continuous inflow feature groups when the algorithm terminates.

Our new algorithm

The FNE-OGSFS can be divided into three parts: online intra-group selection and online interaction analysis, and online redundancy analysis.

Online intra-group selection

Definition 16 Given $\mathrm{OGDS} = (U, G, D, h, t)$, G_t is the feature group arriving at time t, $A = \{A_1, A_2, \cdots, A_r\}$, $A' \subseteq A \subseteq G_t$, if $\mathrm{FNSU}_\delta(A'; D) = \mathrm{FNSU}_\delta(A; D)$, and there exists $\mathrm{FNSU}_\delta(A'; D) > \mathrm{FNSU}_\delta(A' - A_i; D)$ for any $A_i \subseteq A'$, then A' is a reduct of A with respect to D.

Definition 17 Given that $\mathrm{OGDS} = (U, G, D, h, t)$, G_t is the feature group arriving at time t, $A_i \subseteq A$, $A = \{A_1, A_2, \cdots, A_r\} \subseteq G_t$, the significance of feature subset A_i in regard to D is expressed as

$$\mathrm{Sig}(A_i, A; D) = \mathrm{FNSU}_\delta(A; D) - \mathrm{FNSU}_\delta(A - A_i; D). \tag{18}$$

Definition 18 Given $\mathrm{OGDS} = (U, G, D, h, t)$, $A_i \subseteq A$, if $\mathrm{Sig}(A_i, A; D) > 0$, then A_i in A is necessary; otherwise, A_i is unnecessary. If each A_i in A is necessary, then A is independent.

Definition 19 Given $\mathrm{OGDS} = (U, G, D, h, t)$, G_t is the feature group arriving at time t, $A_i \subseteq A$, and $A = \{A_1, A_2, \cdots, A_r\} \subseteq G_t$, if $\mathrm{Sig}(A_i, G_t; D) > 0$, then A_i is a core of G_t.

According to the above theory, a new intra-group streaming feature selection method is demonstrated in Algorithm 1.

Algorithm 1 $\mathrm{FNE} - \mathrm{OGSFS}_{intra}$

Require: $OGDS = (U, G, D, h, t)$ with a new streaming group of features G_t.
Ensure: The selected feature subset S_t'.
1: Initialize $S_t' = \{\}$.
2: **for** each feature f_i in G_t **do**
3: Compute $Sig(f_i, G_t; D) = FNSU_\delta(G_t; D) - FNSU_\delta(G_t - f_i; D)$
4: **if** $Sig(f_i, G_t; D) > 0$ **then**
5: Let $S_t' = S_t' \cup \{f_i\}$
6: **end if**
7: **end for**
8: **return** S_t'

Let the feature group arriving at stamp t be G_t. In Step 3, the significance of each feature in G_t is calculated according to Formula (18), and if the value is greater than 0, it means the feature is important. Then, it will be selected to the feature subset S_t' in Step 5; otherwise, it will be discarded. If all features in G_t are traversed, the algorithm will terminate and return the selected feature subset S_t'.

Online interaction analysis

Definition 20 Given $\mathrm{OGDS} = (U, G, D, h, t)$, $U = x_1, x_2, \ldots, x_n$, G_t is the feature group arriving at time t, $A, B \subseteq G_t$, then the fuzzy neighborhood interaction gain of B with respect to A is denoted as

$$\mathrm{FNIG}_\delta(A, B; D) = FNSU_\delta(A, B; D) - \mathrm{FNSU}_\delta(A; D) - \mathrm{FNSU}_\delta(B; D). \tag{19}$$

Theorem 1 *If* $\mathrm{FNIG}_\delta(A, B; D) > 0$*, then* B *is the interaction feature of* A.

Proof Since $\mathrm{FNIG}_\delta(A, B; D) > 0$, we have that $\mathrm{FNSU}_\delta(A, B; D) > \mathrm{FNSU}_\delta(A; D) + \mathrm{FNSU}_\delta(B; D)$. It is indicated that the information provided by A and B in one piece is more than the sum of the information provided by A and B alone, so A and B interact, and then B is called the interaction feature of A.

Theorem 2 *If* $\mathrm{FNIG}_\delta(A, B; D) \le 0$, *then B is an unnecessary feature of A.*

Proof Since $\mathrm{FNIG}_\delta(A, B; D) \le 0$,we can obtain that $\mathrm{FNSU}_\delta(A, B; D) \le \mathrm{FNSU}_\delta(A; D)+\mathrm{FNSU}_\delta(B; D)$. Thus, the information provided by A and B in one piece is no more than the sum of the information provided by A and B alone. Hence, B is an unnecessary feature of A.

If the newly arrived features in the group are interaction features, further redundancy analysis with the already selected features is needed.

Definition 21 Given OGDS $= (U, G, D, h, t)$, $U = x_1, x_2, \ldots, x_n$, $G_t = \{f_1, f_2, \ldots, f_m\}$ for the feature group arriving at time t, $A \subseteq G_t$, $A = \{A_1, A_2, \cdots, A_r\}$, then the fuzzy neighborhood contrast ratio of A_j with respect to A_i is expressed as

$$FNCR\left(A_i, A_j; D\right) = FNSU_\delta\left(A_j; D\right) - FNSU_\delta\left(A_i; D\right), \tag{20}$$

where $A_i, A_j \in A$ and $i \ne j$.

Theorem 3 *If* $\mathrm{FNCR}\left(A_i, A_j; D\right) \le 0$, *then A_j is the redundant feature of A_i.*

Proof Since $\mathrm{FNCR}\left(A_i, A_j; D\right) \le 0$, from Definition 21, we can have $\mathrm{FNSU}_\delta\left(A_j; D\right) \le \mathrm{FNSU}_\delta(A_i; D)$. It is demonstrated that for two candidate features A_i and A_j, A_i can provide more information that is beneficial for classification; then, A_i is more relevant to D. Therefore, A_j is the redundant feature of A_i.

Theorem 4 *If* $\mathrm{FNCR}\left(A_i, A_j; D\right) > 0$, *then A_j is a relevant feature and A_i is a redundant feature.*

Proof Because $\mathrm{FNCR}\left(A_i, A_j; D\right) > 0$, we can have $\mathrm{FNSU}_\delta\left(A_j; D\right) > \mathrm{FNSU}_\delta(A_i; D)$. Thus, A_j can provide more information that is beneficial for classification; therefore, A_j is more relevant in regard to D, A_j is a relevant feature, and A_i is the redundant feature.

Next, all selected features are re-evaluated by the sparse linear regression model Lasso [75], and the features with similar labels are eliminated based on global group information.

Online redundancy analysis

Given OGDS $= (U, G, D, h, t)$, $U = \{x_1, x_2, \ldots, x_n\}$, let the features that were selected in the above link be $S = \{f_1, f_2, \ldots, f_M\}$, let $X \in R^{M\times n}$ be the dataset matrix and let $\hat{\rho} \in R^M$ be the projection vector. Then, the decision class vector $\hat{y} \in R^n$ is denoted as

$$\hat{y} = X^T\hat{\rho}. \tag{21}$$

Lasso chooses the best $\hat{\rho}$ by minimizing the following objective function:

$$\min_{\hat{\rho}} \| y - X^T\hat{\rho}\|_2 + \gamma \| \hat{\rho}\|_1, \tag{22}$$

where $\|\cdot\|_1$ indicates the L_1 norm of the vector, $\|\cdot\|_2$ indicates the L_2 norm of the vector, and γ is the parameter that regulates the amount of regularization applied to the estimator, whose value is often determined by cross-validation. Lasso can effectively control the number of selected features by setting a part of $\hat{\rho}$ to zero to select features corresponding to nonzero coefficients and adding variable constraints based on the least square method.

Based on all the above investigations, the FNE-OGSFS algorithm is proposed, and the corresponding pseudocode is shown in Algorithm 2. The code is available at https://github.com/SunY-H/OGSFS.

Let the set of features that have been selected at stamp t be S_{t-1} and the set of features selected within the group be S'_t. The fuzzy neighborhood interaction gain of each feature in S'_t relative to S_{t-1} is calculated in Step 3 according to Formula (19) when S'_t flows into the feature space. Based on Theorems 1 and 2, the feature is unnecessary if the interaction gain is not greater than zero; otherwise, the feature is an interaction and needs to be further analyzed for redundancy. In Step 11, the fuzzy neighborhood contrast ratio of each feature in S_{t-1} with respect to the feature is calculated depending on Formula (20). Based on Theorems 3 and 4, the feature is selected into feature subset S if the contrast ratio is greater than zero, while the corresponding redundant features in S_{t-1} are discarded; otherwise, the feature is discarded. If no new feature group flows into the feature space, the FNE-OGSFS algorithm terminates and returns the selected feature subset S after the online redundancy analysis.

Time complexity

For the FNE-OGSFS algorithm, the sample space is $U = \{x_1, x_2, \ldots, x_n\}$, the set of features arriving at time t is $G_t = \left\{f_1, f_2, \ldots, f_{m_t}\right\}^T$, the set of selected features is S_{t-1}, and the set of selected features within the group is S'_t. Each feature in G_t is traversed to calculate its significance in the intra-group feature selection phase, where the computation of

Algorithm 2 FNE − OGSFS

Require: $OGDS = (U, G, D, h, t)$ with fuzzy neighborhood radius δ, group size G.
Ensure: An optimal feature subset S.

Initialize $S = \{\}$.
repeat
  $G_t \leftarrow$ get a new streaming group of features
  /* online intra-group selection */
  $S_t' = \text{FNE} - \text{OGSFS}_{intra}(G_t)$
  /* online interaction analysis */
  for each feature f_i in S_t' **do**
    Compute $FNIG_\delta(S_{t-1}, f_i; D) = FNSU_\delta(S_{t-1}, f_i; D) - FNSU_\delta(S_{t-1}; D) - FNSU_\delta(f_i; D)$
    if $FNIG_\delta(S_{t-1}, f_i; D) > 0$ **then**
      for each feature f_k in S_{t-1} **do**
        Compute $FNCR(f_i, f_k; D) = FNSU_\delta(f_k; D) - FNSU_\delta(f; D)$
        if $FNCR(f_i, f_k; D) \le 0$ **then**
          Break;
        else
          Let $S = S \cup \{f_i\}$; $S = S - \{f_k\}$
        end if
      end for
    end if
  end for
until no more group arrive
/* online redundancy analysis */
$S \leftarrow$ find the global optimal subset by *Lasso* algorithm
return the optimal feature subset S

the parameterized fuzzy neighborhood information granule has the most important impact on the complexity and the time complexity is approximated by $O(m_t n)$. The time complexity of the inter-group feature selection phase increases with the size of the selected feature subset S_{t-1}. Let the number of features in S_{t-1} be $|S_{t-1}|$ and the number of features selected within the group be $|S_t'|$; then, the worst-case time complexity is $O\left(|S_{t-1}| * |S_t'| * n^2\right)$. In addition, the time complexity of the Lasso algorithm is $O(n)$. Therefore, the worst-case time complexity of the FNE-OGSFS algorithm is $O\left(|S_{t-1}| * |S_t'| * n^2\right)$.

Experimental results

The desired effect of our algorithm is to efficiently select a smaller subset of features and obtain a higher classification accuracy. In this section, we conduct a series of experiments based on some existing algorithms and datasets. To provide detailed information regarding the experiments, this section first describes the experimental preparation and then shows the effect of different parameters on the classification performance. By comparing FNE-OGSFS with some popular algorithms, we validate the effectiveness of the proposed algorithms. Finally, we conduct statistical tests on the experimental results.

Table 1 Description of the experimental datasets

No.	Datasets	Samples	Features	Classes
1	Sonar	208	60	2
2	Wpbc	198	34	2
3	Ionosphere	351	33	2
4	Wdbc	569	31	2
5	COLON	62	2000	2
6	DLBCL	77	7129	2
7	LEUKEMIA	72	7129	2
8	LYMPHOMA	62	4026	3
9	SRBCT	83	2308	4
10	Lung Cancer	203	12600	5
11	Ovarian Cancer	253	15154	2
12	MADELON	2600	500	2
13	ARCENE	100	10000	2

Experiment setup

The evaluation framework for feature selection is outlined in Fig. 1. The details of each stage of the experiment are depicted below.

First, the dataset is divided and preprocessed. To verify the feasibility and stability of the developed algorithm, the FNE-OGSFS method is used on thirteen public datasets in our experiments, including four UCI datasets,[1] seven DNA microarray datasets[2] and two NIPS 2003 datasets.[3] All datasets in detail are listed in Table 1. For the missing values in a dataset, such as the LYMPHOMA dataset, the conditional mean completer is used [16]. The 10-fold cross-validation approach is adopted for evaluating the classification performance under different classifiers.

Second, the feature selection methods are selected. In this subsection, FNE-OGSFS is compared with eight state-of-the-art feature selection methods, including two online group streaming feature selection methods (OGSFS-FI [44], Group-SAOLA [42]), three online individual streaming feature selection methods (Alpha-investing [25], SFS-FI [31], OFS-A3M [26]) and three FNRS-based method (FNRS [57], FNCE [67], FNPME-FS [63]). Note that the FNRS-based methods cannot deal with the online feature selection task, and some parameters contained in the above comparison methods need to be specified in advance; here, we refer to the parameter value or value range corresponding to the original thesis.

Finally, the classifier and evaluation metrics are determined. Four classical classifiers, including support vector

[1] http://archive.ics.uci.edu/ml/index.php.
[2] http://csse.szu.edu.cn/staff/zhuzx/.
[3] http://clopinet.com/isabelle/Projects/NIPS2003/.

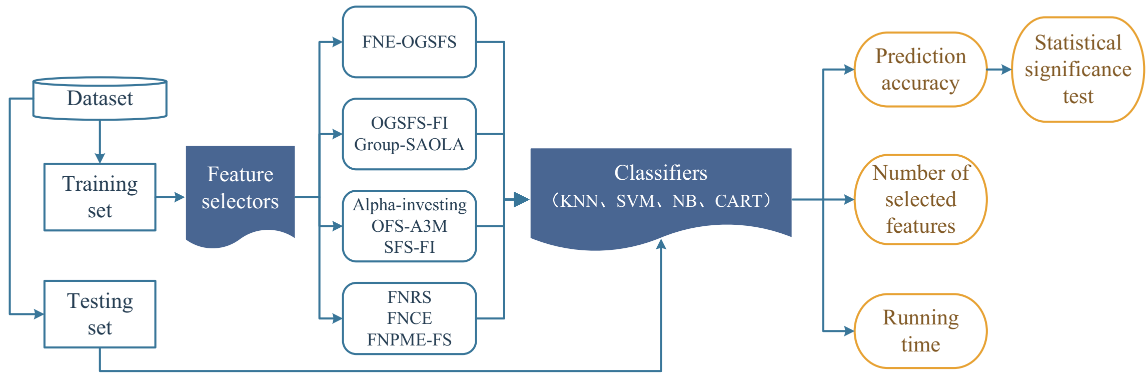


Fig. 1 The evaluation framework for the algorithms

machine (SVM), naive Bayes (NB), K-nearest neighbor (KNN, $k = 3$) and classification and regression tree (CART), are used to evaluate the classification performance of the selected features. The parameters of the classifier are set to the default of MATLAB except for changing the data distribution in the NB classifier to the kernel distribution. We take the predictive accuracy, number of selected features and running time as evaluation metrics of the comparative experiment. Furthermore, the Friedman test and the corresponding post-hoc tests are performed to systematically investigate the statistical performance of the FNE-OGSFS method and its rivals in terms of predictive accuracy.

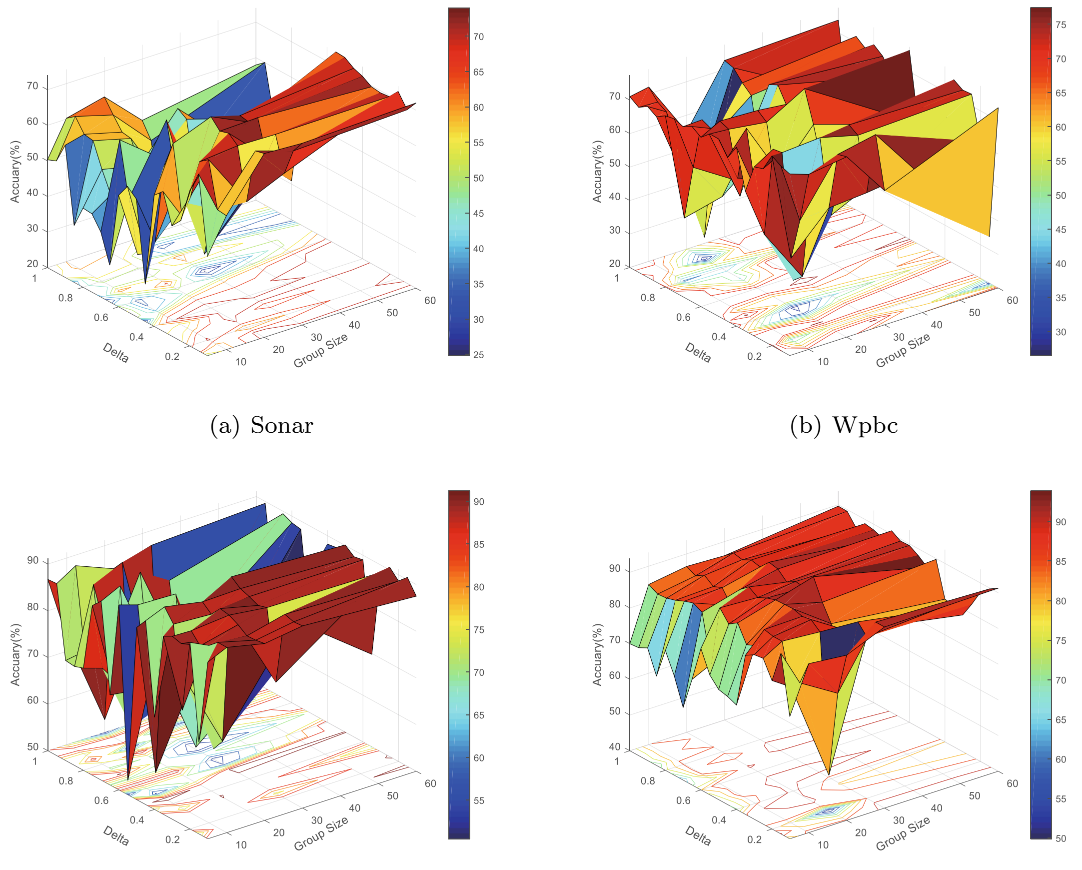
It should be noted that the comparison experiments are based on the same design approach. All experiments are performed in MATLAB R2016a and run in a hardware environment with an Intel Core i5-3470 CPU at 3.20 GHz and 4.0 GB RAM under Windows 10.

The parameter analysis of the FNE-OGSFS

There are two parameters and G in the FNE-OGSFS algorithm. The parameter is used to adjust the size of the fuzzy neighborhood, and the parameter G is applied to control the size of the group. We set the value of from 0.1 to 1 with an interval of 0.05 [60]. Since the dataset does not have a priori group structure information, the experiment obtains the information on the group structure through the artificially specified group size to improve the time efficiency. The values of G are set to 5, 10, 20, 30, and 60 for low-dimensional datasets and 50, 100, 200, 400, and 800 for high-dimensional datasets [44]. In this subsection, we focus on the effect of different parameters on the predictive accuracy, number of selected features and running time.

In terms of predictive accuracy, the variation in predictive accuracy with parameters for thirteen datasets on SVM is shown in Figs. 2 and 3, where four datasets in Fig. 2 are low-dimensional datasets and nine datasets in Fig. 3 are high-dimensional datasets. The experimental results obtained using KNN, NB, and CART are roughly consistent with SVM. Figure 2 indicates that the different parameters have a certain impact on the classification performance of low-dimensional datasets. In detail, the parameter has a deeper influence on some datasets, such as the Sonar and Ionosphere datasets, where the predictive accuracies are generally higher when is less than 0.5. The classification performance of the Wdbc and Wpbc datasets obviously depends more on the parameter G, which can achieve higher predictive accuracies when G is larger, and the influence of is not significant. As seen in Fig. 3, the predictive accuracy on most of the high-dimensional datasets has a significant trend change with parameter changes. The DLBCL, LYMPHOMA and Lung Cancer datasets can achieve better predictive accuracies when both G and are larger. The predictive accuracies of the LEUKEMIA, Ovarian Cancer, ARCENE and MADELON datasets are more influenced by parameter G; when parameter is constant, the predictive accuracy improves significantly with increasing G. The applicability of parameter varies greatly for different datasets, e.g., dataset COLON is more suitable for smaller , and dataset SRBCT is more suitable for larger . Obviously, all datasets can achieve higher predictive accuracy in most regions.

In terms of running time and number of selected features, due to space limitation, three different types of datasets (Wpbc, LEUKEMIA, and ARCENE) are selected as representatives in this subsection to test the experimental effects under different parameters, and the results are shown in Figs. 4 and 5, respectively. Figure 4 shows that the running time increases significantly only when is small in the low-dimensional dataset. The running time and parameter G are closely related in the high-dimensional dataset. This is because as the group size increases, more complex matrix operations occur in the calculation of the fuzzy neighborhood information granule, which in turn consumes more time. Figure 5 indicates that these datasets show better results in terms of compactness (i.e., fewer features are selected in feature selection). Different datasets have different preferences for



(a) Sonar (b) Wpbc

(c) Ionosphere (d) Wdbc

Fig. 2 The predictive accuracy varying with different values of δ and G on four low-dimensional datasets

the parameter δ. Fewer features are selected when G is small because fewer relevant features are considered calculating the fuzzy neighborhood information granule; hence, more features are discarded in the intra-group feature selection.

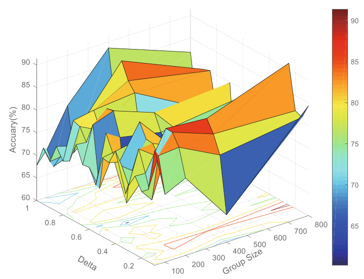
Overall, the experimental result demonstrates the effectiveness of FNE-OGSFS in selecting the optimal feature subsets for different types of datasets. It should be noted that the parameters corresponding to the best predictive accuracy are different for the thirteen datasets. Therefore, the parameters need to be determined in advance before feature selection for the datasets to achieve the maximum balance among higher accuracy, smaller running time, and greater compactness.

Comparison with other algorithms

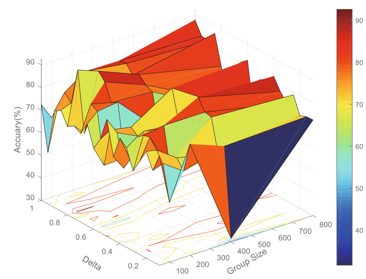
In this subsection, the performance of FNE-OGSFS and its rivals in terms of the predictive accuracy, number of selected features and running time are analyzed.

Tables 2, 3, 4 and 5 show the predictive accuracies of the KNN, SVM, NB, and CART classifiers. The last two rows list the win/tie/lose (abbreviated as W/T/L) counts and the average predictive accuracies of the algorithms on all datasets, with bold font indicating the highest predictive accuracy. Tables 6 and 7 show the number of selected features and running times of the nine algorithms, respectively. Specifically, we discuss the following.

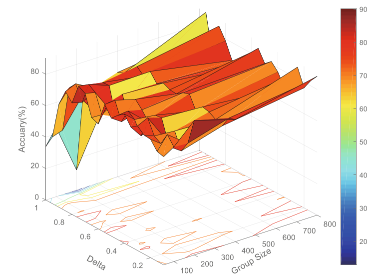
To more intuitively confirm the algorithm effectiveness, we plotted spider web graphs to depict the average predictive accuracy on each classifier, as shown in Fig. 6, where the red line represents the predictive accuracy of our proposed algorithm on each dataset. Tables 2, 3, 4 and 5 and Fig. 6 show that FNE-OGSFS performs significantly better than the other comparison algorithms in terms of overall classification performance. The average predictive accuracy reaches the maximum on all classifiers, and the win counts achieve the highest among all comparison algorithms. By intra-group feature selection, our proposed algorithm select features with high significance. During the online interaction analysis, FNE-OGSFS leave the features with interaction. The experimental results show that this strategy is effective and the selected features can achieve high prediction accuracy. Com-



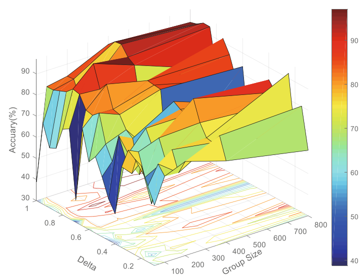
(a) COLON



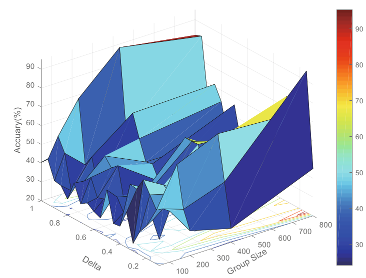
(b) DLBCL



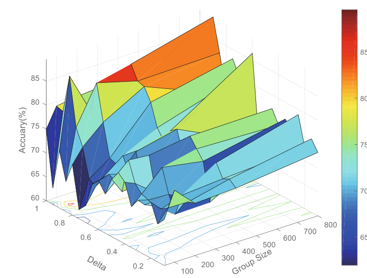
(c) LEUKEMIA



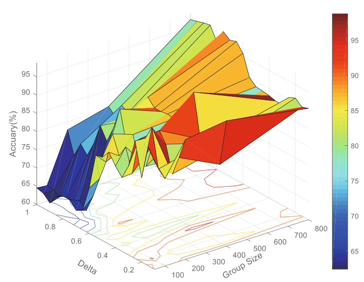
(d) LYMPHOMA



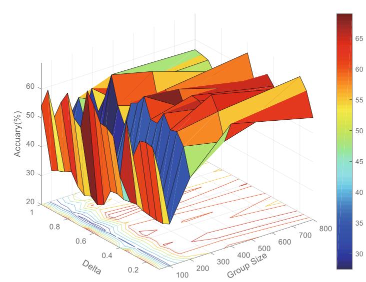
(e) SRBCT



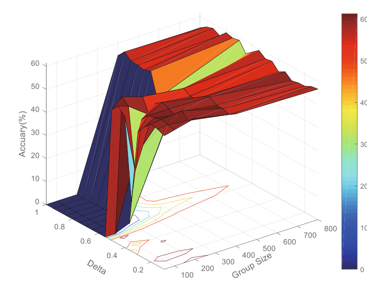
(f) Lung Cancer



(g) Ovarian Cancer

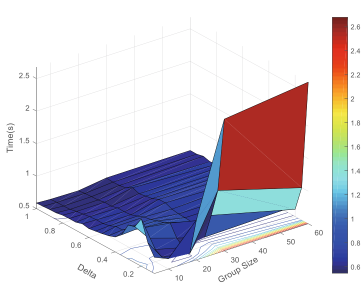


(h) ARCENE

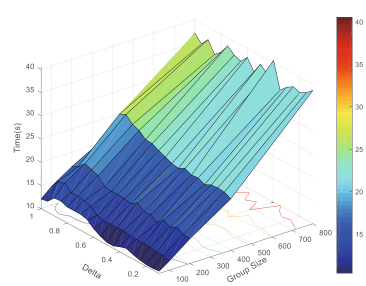


(i) MADELON

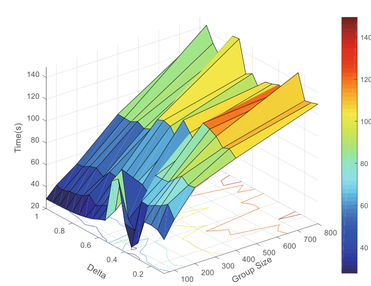
Fig. 3 Predictive accuracy varying with different values of δ and G on nine high-dimensional datasets



(a) Wdbc



(b) LEUKEMIA



(c) MADELON

Fig. 4 Running time varying with different values of δ and G on three representative datasets

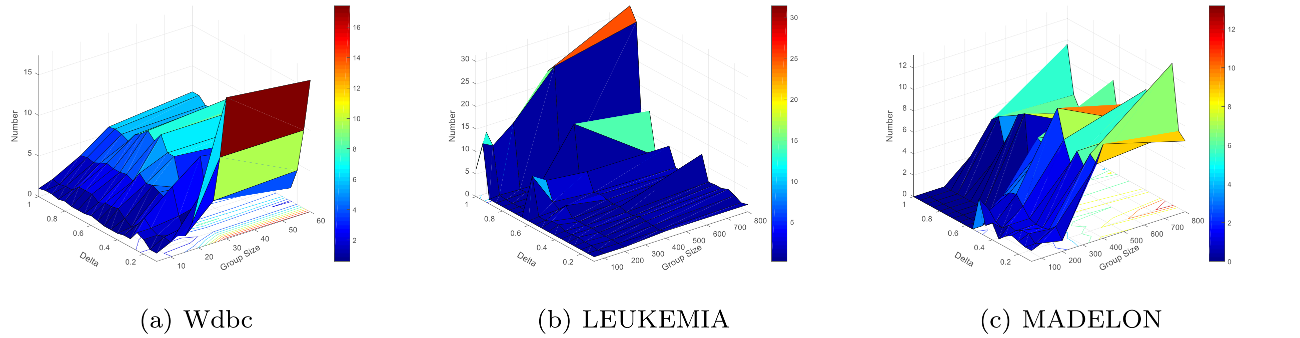


Fig. 5 Number of selected features varying with different values of δ and G on three representative datasets

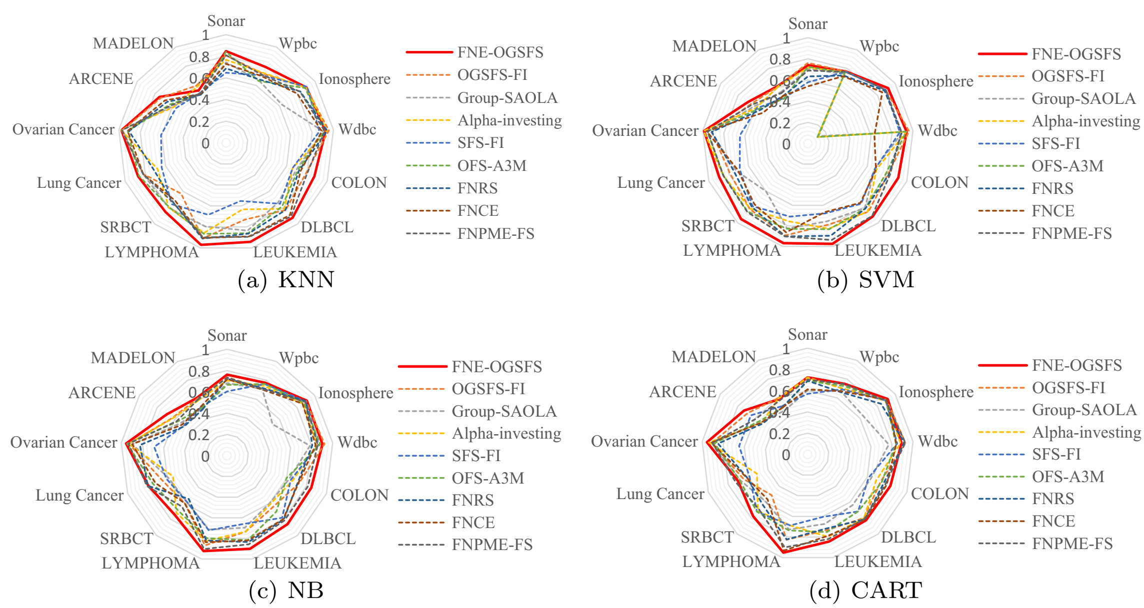


Fig. 6 Spider web diagram showing the predictive accuracy of thirteen datasets on four classifiers

pared with online streaming feature selection methods, the FNE-OGSFS method performs significantly better on genetic datasets. Because the algorithm can handle the fuzziness and uncertainty of genetic datasets well by using the uncertainty measures based on fuzzy neighborhood symmetric uncertainty. Compared with the FNRS-based feature selection methods, our algorithm has a significant advantage because our algorithm incorporates the coincidence degree that allows the selection of highly separability features, which is beneficial for classification.

Compared with the number of selected features, we find that the proposed algorithm can achieve feature reduction. In the last part of FNE-OGSFS, namely, online redundancy analysis, redundant features can be effectively eliminated, which is helpful in selecting fewer features. Although Group-SAOLA and SFS-FI select fewer features, they are far less accurate than our proposed algorithm and other comparison algorithms. The features removed in the redundancy analysis phase of our algorithm also do not degrade the classification performance.

By comparing the time used by the algorithms, FNE-OGSFS can achieve high efficiency when dealing with low-dimensional datasets. However, its performance is poor when dealing with high-dimensional datasets such as the Lung Cancer and Ovarian Cancer datasets because the process of computing fuzzy neighborhood granules consumes considerable time. This situation is more evident in all three comparison algorithms based on FNRS. Our algorithm requires re-evaluation of the selected features and thus runs slower. However, since we evaluate the interactivity and redundancy of the selected features, we select more favorable and fewer features for classification, which performs better in terms of compactness and accuracy.

In conclusion, FNE-OGSFS provides the best overall performance on four classical classifiers. Although FNE-OGSFS has a slightly longer running time, it achieves the

Table 2 Comparison of predictive accuracies on classifier KNN

Datasets	FNE- OGSFS	OGSFS- FI	Group- SAOLA	Alpha- investing	SFS- FI	OFS- A3M	FNRS	FNCE	FNPME- FS
Sonar	**0.8478**	0.8091	0.744	0.7689	0.6487	0.8419	0.6981	0.7312	0.8102
Wpbc	**0.7912**	0.723	0.6522	0.7323	0.7041	0.6457	0.6613	0.6986	0.7127
Ionosphere	0.9019	0.8977	0.6254	0.8994	**0.9115**	0.8906	0.8342	0.8011	0.8311
Wdbc	0.9274	**0.9579**	0.9013	0.939	0.8696	0.9137	0.942	0.8912	0.9162
COLON	**0.8734**	0.7905	0.6966	0.657	0.6572	0.7105	0.7214	0.7365	0.8014
DLBCL	**0.9267**	0.8298	0.7306	0.8121	0.7527	0.7816	0.828	0.8901	0.9069
LEUKEMIA	**0.9417**	0.729	0.834	0.6332	0.5519	0.8595	0.8639	0.8895	0.8944
LYMPHOMA	**0.9703**	0.9085	0.7963	0.858	0.6852	0.8651	0.911	0.9039	0.8991
SRBCT	**0.8495**	0.6273	0.8013	0.8005	0.7585	0.7944	0.7367	0.7121	0.738
Lung Cancer	**0.8699**	0.8173	0.8197	0.6779	0.6375	0.8621	0.7132	0.8196	0.8203
Ovarian Cancer	0.9746	0.9652	0.9742	**0.9791**	0.6078	0.9166	0.9044	0.9162	0.9718
ARCENE	**0.7479**	0.733	0.5967	0.5848	0.5635	0.628	0.6649	0.6895	0.6124
MADELON	0.548	**0.5976**	0.509	0.5836	0.521	0.511	0.503	0.5091	0.5132
W/T/L	9/0/4	2/0/11	0/0/13	1/0/12	1/0/12	0/0/13	0/0/13	0/0/13	0/0/13
Average	**0.8593**	0.7989	0.7447	0.7635	0.6822	0.7862	0.7672	0.7837	0.8021

Table 3 Comparison of predictive accuracies on classifier SVM

Datasets	FNE- OGSFS	OGSFS- FI	Group- SAOLA	Alpha- investing	SFS- FI	OFS- A3M	FNRS	FNCE	FNPME- FS
Sonar	0.74	**0.7626**	0.707	0.7168	0.5778	0.7256	0.6337	0.5413	0.6891
Wpbc	**0.7691**	0.7687	0.7426	0.7631	0.7658	0.7538	0.7262	0.7174	0.7672
Ionosphere	**0.922**	0.8926	0.1226	0.1083	0.8922	0.1103	0.8791	0.853	0.8932
Wdbc	0.9385	**0.9634**	0.9026	0.9372	0.8752	0.9342	0.891	0.6316	0.8897
COLON	**0.9138**	0.8095	0.674	0.6926	0.6759	0.7899	0.8203	0.6869	0.8292
DLBCL	**0.9201**	0.8608	0.8115	0.8543	0.7589	0.8183	0.8177	0.7473	0.9136
LEUKEMIA	**0.9746**	0.7896	0.7563	0.8281	0.6798	0.8329	0.896	0.6653	0.938
LYMPHOMA	**0.9679**	0.8939	0.8263	0.755	0.7105	0.8279	0.9036	0.8642	0.9021
SRBCT	**0.9526**	0.8267	0.5989	0.8201	0.7899	0.8575	0.7991	0.7633	0.8629
Lung Cancer	**0.8955**	0.788	0.7244	0.8557	0.6897	0.8524	0.746	0.7061	0.8542
Ovarian Cancer	0.9877	0.9141	0.921	**0.9927**	0.6463	0.9579	0.9152	0.9637	0.9536
ARCENE	**0.69**	0.6553	0.6181	0.5928	0.587	0.5714	0.5531	0.513	0.6455
MADELON	**0.6133**	0.6091	0.5473	0.607	0.4849	0.4795	0.5063	0.5064	0.5389
W/T/L	10/0/3	2/0/11	0/0/13	1/0/12	0/0/13	0/0/13	0/0/13	0/0/13	0/0/13
Average	**0.8681**	0.8103	0.6887	0.7326	0.7026	0.7317	0.7759	0.7046	0.8213

Table 4 Comparison of predictive accuracies on classifier NB

Datasets	FNE- OGSFS	OGSFS- FI	Group- SAOLA	Alpha- investing	SFS- FI	OFS- A3M	FNRS	FNCE	FNPME- FS
Sonar	**0.7632**	0.7389	0.684	0.7078	0.6039	0.6673	0.7233	0.7192	0.7338
Wpbc	**0.7742**	0.7152	0.7105	0.7421	0.7616	0.7623	0.719	0.7039	0.7093
Ionosphere	**0.9113**	0.9027	0.5196	0.8624	0.8984	0.8766	0.9037	0.8632	0.8932
Wdbc	0.904	**0.9223**	0.786	0.8539	0.8668	0.8821	0.8075	0.8395	0.858
COLON	**0.8481**	0.7238	0.6215	0.6594	0.6405	0.5965	0.7112	0.7181	0.8127
DLBCL	**0.8577**	0.6477	0.6011	0.6096	0.7799	0.7006	0.8019	0.7956	0.8131
LEUKEMIA	**0.9011**	0.7336	0.6973	0.7347	0.6591	0.8173	0.8346	0.8198	0.8571
LYMPHOMA	**0.9223**	0.8681	0.7139	0.8251	0.721	0.7981	0.8199	0.8366	0.9037
SRBCT	**0.742**	0.5983	0.5732	0.7176	0.5732	0.6898	0.5422	0.5981	0.6392
Lung Cancer	0.7925	0.6508	0.5892	0.5432	0.6107	0.7433	0.7972	0.7211	**0.806**
Ovarian Cancer	**0.956**	0.9298	0.8971	0.9333	0.6903	0.9468	0.8261	0.8973	0.9318
ARCENE	**0.6874**	0.625	0.5357	0.6191	0.4961	0.565	0.488	0.5144	0.5742
MADELON	**0.6131**	0.6048	0.5011	0.587	0.5194	0.5075	0.4931	0.5239	0.594
W/T/L	11/0/2	1/0/12	0/0/13	0/0/13	0/0/13	0/0/13	0/0/13	0/0/13	1/0/12
Average	**0.821**	0.7432	0.6485	0.7227	0.6785	0.7349	0.7283	0.7347	0.7789

Table 5 Comparison of predictive accuracies on classifier CART

Datasets	FNE- OGSFS	OGSFS- FI	Group- SAOLA	Alpha- investing	SFS- FI	OFS- A3M	FNRS	FNCE	FNPME- FS
Sonar	0.7253	**0.7283**	0.6018	0.7192	0.5715	0.6918	0.6946	0.6139	0.7046
Wpbc	**0.7492**	0.7012	0.6735	0.7012	0.692	0.7288	0.6633	0.6722	0.7481
Ionosphere	**0.9168**	0.9053	0.6362	0.8897	0.8938	0.8947	0.8413	0.9075	0.8912
Wdbc	0.8895	0.9266	0.7692	0.8817	0.8488	0.8374	0.916	0.8403	**0.927**
COLON	**0.8338**	0.7327	0.6385	0.6897	0.6153	0.7452	0.7655	0.7962	0.7836
DLBCL	**0.8275**	0.798	0.6373	0.7993	0.7157	0.7163	0.8132	0.8199	0.8081
LEUKEMIA	**0.8439**	0.7352	0.6721	0.7978	0.6051	0.7527	0.719	0.7894	0.8205
LYMPHOMA	**0.9513**	0.8267	0.7276	0.6884	0.6895	0.7363	0.8239	0.9316	0.9023
SRBCT	**0.7753**	0.5108	0.5949	0.7069	0.7184	0.7192	0.6348	0.5672	0.6911
Lung Cancer	0.7007	0.6477	0.6805	0.5091	0.5845	0.6853	0.6099	0.615	**0.7265**
Ovarian Cancer	**0.9538**	0.9075	0.9122	0.9379	0.6504	0.8844	0.8502	0.9134	0.9027
ARCENE	**0.7267**	0.6845	0.5321	0.526	0.6492	0.5507	0.5116	0.5268	0.5912
MADELON	0.5863	**0.5977**	0.466	0.5949	0.5155	0.5005	0.5047	0.5014	0.573
W/T/L	9/0/4	2/0/11	0/0/13	0/0/13	0/0/13	0/0/13	0/0/13	0/0/13	2/0/11
Average	**0.8062**	0.7463	0.6571	0.7263	0.6731	0.7264	0.7191	0.7303	0.7746

Table 6 Comparison of number of selected features

Datasets	FNE- OGSFS	OGSFS- FI	Group- SAOLA	Alpha- investing	SFS- FI	OFS- A3M	FNRS	FNCE	FNPME- FS
Sonar	8.8	12.4	1	15.4	1	21.8	1.8	1.2	10.2
Wpbc	1.6	2.2	2.6	8	1	1	9	11.4	8.6
Ionosphere	5	3.4	1	16.2	2	8.8	4.2	3.8	4
Wdbc	6.6	11.6	10.2	17.2	1	16	12.6	10.2	8.2
COLON	5.8	6.2	1	2.2	1	31.6	3	7.2	11.6
DLBCL	5.2	11.2	8.6	6.2	1.4	29.2	21.4	10.6	8.6
LEUKEMIA	8.6	10	2.2	7.6	1.4	23.4	46.6	12.2	10
LYMPHOMA	7.4	10.6	6.2	20.6	2.6	30.6	32.6	21.4	14.6
SRBCT	24.8	6	6.6	18.8	72	13.2	38.6	19.2	15.2
Lung Cancer	14.2	11.8	12.2	30.6	1	40.2	84.2	38.2	50.6
Ovarian Cancer	6.4	47.8	21	51.4	1	7.6	62.4	26.6	17.8
ARCENE	1.4	22.6	10.4	3	1	30.8	4.8	1.8	23.2
MADELON	4.6	9.8	2.6	6.8	1	2	1	2.6	10.2
Average	7.7	12.7	6.6	15.7	6.7	19.7	24.8	12.8	14.8

Table 7 Comparison of running times(s)

Datasets	FNE- OGSFS	OGSFS- FI	Group- SAOLA	Alpha- investing	SFS- FI	OFS- A3M	FNRS	FNCE	FNPME- FS
Sonar	0.3386	0.0237	0.0091	0.0024	0.0015	0.1673	0.3966	0.8133	0.4657
Wpbc	0.159	0.0296	0.0125	0.0011	0.0017	0.0439	0.2032	0.3906	0.2546
Ionosphere	0.3941	0.052	0.0166	0.0153	0.0012	0.1938	0.3619	1.2767	0.4993
Wdbc	0.7986	0.5163	0.0089	0.0084	0.0011	0.6186	0.7101	2.8849	0.9722
COLON	3.5324	0.2213	0.1689	0.0606	0.0467	0.9483	5.228	11.7071	7.2214
DLBCL	9.4213	0.6451	0.6014	0.3338	0.1957	5.2565	8.3612	9.7326	9.9175
LEUKEMIA	11.0722	0.524	0.7435	0.4950	0.2844	6.4142	34.5264	26.5811	30.7039
LYMPHOMA	9.1881	0.2439	0.6631	0.4456	0.1241	15.2325	12.3147	14.275	18.3367
SRBCT	11.5211	0.5226	0.2154	0.0941	0.9237	3.0298	15.039	15.7192	16.9631
Lung Cancer	184.463	0.6432	3.4635	1.871	0.7521	65.0008	255.4301	271.3386	307.2651
Ovarian Cancer	144.7541	2.8841	4.7695	3.4376	1.0179	66.4503	202.4256	236.786	210.9788
ARCENE	36.81	5.1314	2.0133	0.9418	0.5137	21.5113	40.2291	52.3619	57.1783
MADELON	94.5696	21.0849	0.0862	0.0975	0.0574	146.9587	71.3425	79.0516	89.4967
Average	39.0017	2.5017	0.9825	0.6003	0.3016	25.5251	49.736	55.6091	57.7118

highest average prediction accuracy and effectively removes redundant features. It has been verified that FNE-OGSFS is superior to the other compared algorithms on different types of datasets.

Statistical significance analysis

To further explore the generalization ability of the FNE-OGSFS algorithm systematically, the Friedman test and Nemenyi post-hoc test are performed in this subsection to assess the statistical significance of the algorithm [76]. The average predictive accuracies of each algorithm on the thirteen datasets shown in Tables 4, 5, 6 and 7 are ranked from lowest to highest before using the Friedman test, and the rankings are divided equally when the performance is the same. The Friedman statistic is described as

$$\chi_F^2 = \frac{12n}{h\,(h+1)}\left(\sum_{j=1}^{h} R_j^2 - \frac{h(h+1)^2}{4}\right), \tag{23}$$

$$F_F = \frac{(n-1)\,\chi_F^2}{n\,(h-1) - \chi_F^2}, \tag{24}$$

where n and h are the number of datasets and algorithms, respectively, and $R_j\,(j = 1, 2, \ldots, h)$ denotes the average ranking of the j th algorithm over all datasets. The variable χ_F^2 obeys the χ^2 distribution with $h-1$ degrees of freedom, and the variable F_F obeys the F distribution with $h-1$ and $(h-1)\,(n-1)$ degrees of freedom. To further obtain the difference between the algorithms, the critical difference (CD) of the mean rankings in the Nemenyi test is calculated by

$$CD_\alpha = q_\alpha \sqrt{\frac{h\,(h+1)}{6n}}, \tag{25}$$

where α represents the significance level of the Nemenyi test and q_α represents the critical value corresponding to the number of comparison algorithms at a particular significance level.

The average rankings of predictive accuracy of a particular algorithm on all datasets were acquired according to the statistical tests provided in [76]. For the predictive accuracy on thirteen datasets in Tables 2, 3, 4 and 5, the Friedman tests were achieved by the comparison of FNE-OGSFS with the other algorithms. The null hypothesis of the Friedman test is established when all algorithms are equal in metrics of predictive accuracy. Table 8 describes the average ranking of the nine algorithms and the values of χ_F^2 and F_F on the four classifiers.

The F_F distribution has 8 and 96 degrees of freedom when $n = 13$ and $h = 9$, respectively. By checking the table, we can obtain the value of χ_F^2 (8) in the χ_F^2 distribution as 13.36

Table 8 Statistical test of nine methods on four classifiers

Classifiers	Mean rankings									χ_F^2	F_F
	FNE- OGSFS	OGSFS- FI	Group- SAOLA	Alpha- investing	SFS- FI	OFS- A3M	FNRS	FNCE	FNPME- FS		
KNN	1.54	4	6.46	5	7.23	5.31	5.69	5.62	4.15	37.71	6.83
SVM	1.23	3.46	6.46	4.62	7.38	5.23	5.69	7.46	3.46	57.85	15.04
NB	1.23	4	7.85	5.38	6.69	5.15	5.62	5.62	3.46	51.13	11.61
CART	1.54	4.15	7.23	5.23	6.69	5.31	6.08	4.92	3.46	35.22	6.14

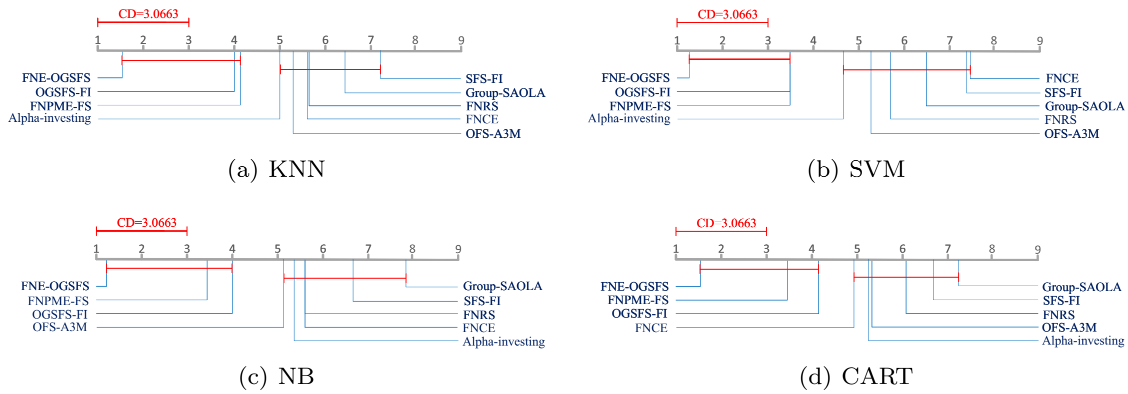


Fig. 7 The test of Nemenyi on four classifiers

and $F(8, 96)$ in the F distribution as 1.74 when the significance level is $= 0.1$. From the results in Table 8, we can see that the values of χ_F^2 and F_F on four classifiers are greater than their values on $\chi_F^2(8)$ and $F(8, 96)$. That is, all null hypotheses are rejected, which demonstrates that the performances of these algorithms are significantly different. Further post-hoc tests are performed next to obtain the difference between the algorithms. The critical value $q_{0.1}$ is 2.855 when $h = 9$ and then the critical range $CD_{0.1}$ is 3.0663. To more intuitively compare the differences of the algorithms, a graph is introduced to connect the methods that do not differ significantly from each other, in which the critical values among all algorithms can be clearly illustrated. Fig. 7 shows the comparison of FNE-OGSFS with the other algorithms on four classifiers, where the critical value and its range are shown above the axis, the coordinate axis plots the average ranking values for each algorithm, and the average ranking of the left-hand side is the lowest. The horizontal lines are used to connect the algorithms with no significant difference, which indicates that any two algorithms with a difference in average ranking less than the value of CD are connected by the red line.

As shown in Fig. 7, the significant difference among the nine algorithms are obvious. FNE-OGSFS performs significantly better than the other algorithms on four classifiers. In some cases, the significance of the algorithms on different classifiers is slightly different, e.g., Group-SAOLA has the lowest average ranking on NB and CART classifiers, but not on the other classifiers. FNE-OGSFS has the same group as FNPME-FS and OGSFS-FI, which means the differences among the three algorithms are not obvious. However, the difference in their average rankings is very close to the critical value on most classifiers, so it can still be concluded that FNE-OGSFS excels against the two compared methods. In summary, FNE-OGSFS outperforms the other eight compared algorithms overall.

Conclusion and future work

In this paper, we proposed a novel online group streaming feature selection method, named FNE-OGSFS. First, a new separability measure was investigated, and some fuzzy neighborhood entropy-based uncertainty measures were expanded, inspired by both algebra and information views. Second, intra-group feature selection was performed according to the significance of features. Then, interactive feature selection was devised in an online manner for features that flow into the feature space. Finally, the Lasso model was applied to online redundancy analysis. Compared to some state-of-the-art online streaming feature selection methods and traditional feature selection methods based on FNRS, FNE-OGSFS demonstrated better comprehensive performance.

In future work, we will further optimize the method, focusing on how to select the best parameters automatically and improve the computational efficiency of the algorithm and achieve an optimal balance on high-dimensional datasets. Moreover, research on incremental feature selection with feature streams and data streams based on fuzzy neighborhood rough sets will receive more attention.

Acknowledgements This work was supported by the National Natural Science Foundation of China under Grant (6197 6082, 62076089, 62002103).

References

1. Shen C-N, Zhang K (2021) Two-stage improved Grey Wolf optimization algorithm for feature selection on high-dimensional classification. Complex Intell Syst. https://doi.org/10.1007/s40747-021-00452-4
2. Sun L, Zhang X-Y, Qian Y-H, Xu J-C, Zhang S-G (2019) Feature selection using neighborhood entropy-based uncertainty measures for gene expression data classification. Inf Sci 502:18–41. https://doi.org/10.1016/j.ins.2019.05.072
3. Wang C-Z, Huang Y, Shao M-W, Hu Q-H, Chen D-G (2020) Feature selection based on neighborhood self-information. IEEE Trans Cybern 50(9):4031–4042. https://doi.org/10.1109/TCYB.2019.2923430
4. Sang B-B, Chen H-M, Yang L, Li T-R, Xu W-H, Luo C (2021) Feature selection for dynamic interval-valued ordered data based on fuzzy dominance neighborhood rough set. Knowl Based Syst 227:107223. https://doi.org/10.1016/j.knosys.2021.107223
5. Shen H-T, Zhu Y-H, Zheng W, Zhu X-F (2021) Half-quadratic minimization for unsupervised feature selection on incomplete data. IEEE Trans Neural Netw Learn Syst 32(7):3122–3135. https://doi.org/10.1109/TNNLS.2020.3009632
6. Onar SC, Oztaysi B, Kahraman C (2017) Dynamic intuitionistic fuzzy multi-attribute aftersales performance evaluation. Complex Intell Syst 3(3):197–204. https://doi.org/10.1007/s40747-017-0047-7
7. Ding W, Stepinski TF, Mu Y, Bandeira L, Ricardo R, Wu Y-X, Lu Z-Y, Cao T-Y, Wu X-D (2011) Subkilometer crater discovery with boosting and transfer learning. ACM Trans Intell Syst Technol 2(4):1–22. https://doi.org/10.1145/1989734.1989743
8. Bai S-X, Lin Y-J, Lv Y, Chen J-K, Wang C-X (2021) Kernelized fuzzy rough sets based online streaming feature selection for large-scale hierarchical classification. Appl Intell 51(3):1602–1615. https://doi.org/10.1007/s10489-020-01863-5
9. Yu K, Ding W, Wu X-D (2016) LOFS: a library of online streaming feature selection. Knowl Based Syst 113:1–3. https://doi.org/10.1016/j.knosys.2016.08.026
10. Lv Y, Lin Y-J, Chen X-Y, Wang D-X, Wang C-X (2020) Online streaming feature selection based on feature interaction. In: 2020 IEEE international conference on knowledge graph. https://doi.org/10.1109/ICBK50248.2020.00017
11. Wang J-L, Zhao P-L, Hoi SCH, Jin R (2013) Online feature selection and its applications. IEEE Trans Knowl Data Eng 26(3):698–710. https://doi.org/10.1109/TKDE.2013.32
12. Wu D, He Y, Luo X, Zhou M-C (2021) A latent factor analysis-based approach to online sparse streaming feature selection. IEEE Trans Syst Man Cybern. https://doi.org/10.1109/TSMC.2021.3096065
13. Lu J, Liu A-J, Song Y-L, Zhang G-Q (2020) Data-driven decision support under concept drift in streamed big data. Complex Intell Syst 6(1):157–163. https://doi.org/10.1007/s40747-019-00124-4
14. Wang Y, Ding Y, He X-J, Fan X, Lin C, Li F-Q, Wang T-Z, Luo Z-X, Luo J-B (2021) Novelty detection and online learning for chunk data streams. IEEE Trans Pattern Anal Mach Intell 43(7):2400–2412. https://doi.org/10.1109/TPAMI.2020.2965531
15. Diao R, Parthalain NM, Shen Q (2013) Dynamic feature selection with fuzzy-rough sets. In: 2013 IEEE international conference on fuzzy systems. https://doi.org/10.1109/FUZZ-IEEE.2013.6622410
16. Wan J-H, Chen H-M, Li T-R, Yang X-L, Sang B-B (2021) Dynamic interaction feature selection based on fuzzy rough set. Inf Sci 581:891–911. https://doi.org/10.1016/j.ins.2021.10.026
17. Sangma JW, Sarkar M, Pal V, Agrawal A, Yogita (2022) Hierarchical clustering for multiple nominal data streams with evolving behaviour. Complex Intell Syst. https://doi.org/10.1007/s40747-021-00634-0
18. Yang S-Y, Zhang H-Y, De Baets B, Jah M, Shi G (2021) Quantitative dominance-based neighborhood rough sets via fuzzy preference relations. IEEE Trans Fuzzy Syst 29(3):515–529. https://doi.org/10.1109/TFUZZ.2019.2955883
19. Li S-J, Zhang K-X, Li Y-L, Wang S-Q, Zhang S-Q (2021) Online streaming feature selection based on neighborhood rough set. Appl Soft Comput 113(A):108025. https://doi.org/10.1016/j.asoc.2021.108025
20. Wang H-M, Wang G-Y, Zeng X-H, Peng S-Y (2017) Online streaming feature selection based on conditional information entropy. In:2017 IEEE international conference on big knowledge. https://doi.org/10.1109/ICBK.2017.44
21. Paul D, Jain A, Saha S, Mathew J (2021) Multi objective PSO based online feature selection for multi label classification. Knowl Based Syst 222:106966. https://doi.org/10.1016/j.knosys.2021.106966
22. Zhou P, Hu X-G, Li P-P, Wu X-D (2019) OFS-density: a novel online streaming feature selection method. Pattern Recognit 86:48–61. https://doi.org/10.1016/j.patcog.2018.08.009
23. Zhou P, Hu X-G, Li P-P, Wu X-D (2017) Online feature selection for high-dimensional class-imbalanced data. Knowl Based Syst 136:187–199. https://doi.org/10.1016/j.knosys.2017.09.006
24. Liu J-H, Lin Y-J, Li Y-W, Weng W, Wu S-X (2018) Online multi-label streaming feature selection based on neighborhood rough set. Pattern Recognit 84:273–287. https://doi.org/10.1016/j.patcog.2018.07.021
25. Zhou J, Foster DP, Stine RA, Ungar LH (2006) Streamwise feature selection. J Mach Learn Res 7:1861–1885
26. Zhou P, Hu X-G, Li P-P, Wu X-D (2019) Online streaming feature selection using adapted neighborhood rough set. Inf Sci 481:258–279. https://doi.org/10.1016/j.ins.2018.12.074
27. Yu K, Wu X-D, Ding W, Pei J (2014) Towards scalable and accurate online feature selection for big data. In:2014 IEEE international conference on data mining. https://doi.org/10.1109/ICDM.2014.63
28. Rahmaninia M, Moradi P (2018) OSFSMI: online stream feature selection method based on mutual information. Appl Soft Comput 68:733–746. https://doi.org/10.1016/j.asoc.2017.08.034
29. Perkins S, Theiler J (2003) Online feature selection using grafting. In: Proceedings of the twentieth international conference on international conference on machine learning, pp 592–599
30. Wu X-D, Yu K, Ding W, Wang H, Zhu X-Q (2013) Online feature selection with streaming features. IEEE Trans Pattern Anal Mach Intell 35(5):1178–1192. https://doi.org/10.1109/TPAMI.2012.197
31. Zhou P, Li P-P, Zhao S, Wu X-D (2021) Feature interaction for streaming feature selection. IEEE Trans Neural Netw Learn Syst 32(10):4691–4702. https://doi.org/10.1109/TNNLS.2020.3025922
32. Eskandari S, Javidi MM (2016) Online streaming feature selection using rough sets. Int J Approx Reason 69:35–57. https://doi.org/10.1016/j.ijar.2015.11.006
33. Javidi MM, Eskandari S (2019) Online streaming feature selection: a minimum redundancy, maximum significance approach. Pattern Anal Appl 22(3):949–963. https://doi.org/10.1007/s10044-018-0690-7
34. Lin Y-J, Hu Q-H, Liu J-H, Li J-J, Wu X-D (2017) Streaming feature selection for multilabel learning based on fuzzy mutual information. IEEE Trans Fuzzy Syst 25(6):1491–1507. https://doi.org/10.1109/TFUZZ.2017.2735947
35. You D-L, Li R-Q, Liang S-P, Sun M-M, Ou X-J, Yuan F-Y, Shen L-M, Wu X-D (2021) Online causal feature selection for streaming features. IEEE Trans Neural Netw Learn Syst. https://doi.org/10.1109/TNNLS.2021.3105585
36. Yang L, Qin K-Y, Sang B-B, Xu W-H (2021) Dynamic fuzzy neighborhood rough set approach for interval-valued informa-

tion systems with fuzzy decision. Appl Soft Comput 111:107679. https://doi.org/10.1016/j.asoc.2021.107679
37. Basiri ME, Abdar M, Cifci MA, Nemati S, Acharya UR (2020) A novel method for sentiment classification of drug reviews using fusion of deep and machine learning techniques. Knowl Based Syst 198:105949. https://doi.org/10.1016/j.knosys.2020.105949
38. Wang M, Li H, Tao D-C, Lu K, Wu X-D (2012) Multimodal graph based ranking for web image search. IEEE Trans Image Process 21(11):4649–4661. https://doi.org/10.1109/TIP.2012.2207397
39. Al Nuaimi N, Masud MM (2020) Online streaming feature selection with incremental feature grouping. Wires Data Min Knowl 10(4):e1364. https://doi.org/10.1002/widm.1364
40. Li H-G, Wu X-D, Li Z, Ding W (2014) Group feature selection with streaming features. In: 2013 IEEE 13th international conference on data mining. https://doi.org/10.1109/ICDM.2013.137
41. Wang J, Wang M, Li P-P, Liu L-Q, Zhao Z-Q, Hu X-G, Wu X-D (2015) Online feature selection with group structure analysis. IEEE Trans Knowl Data Eng 27(11):3029–3041. https://doi.org/10.1109/TKDE.2015.2441716
42. Yu K, Wu X-D, Ding W, Pei J (2016) Scalable and accurate online feature selection for big data. ACM Trans Knowl Discov Data 11(2):1–39. https://doi.org/10.1145/2976744
43. Liu J-H, Lin Y-J, Wu S-X, Wang C-X (2018) Online multi-label group feature selection. Knowl Based Syst 143:42–57. https://doi.org/10.1016/j.knosys.2017.12.008
44. Zhou P, Wang N, Zhao S (2021) Online group streaming feature selection considering feature interaction. Knowl Based Syst 226:107157. https://doi.org/10.1016/j.knosys.2021.107157
45. Zhang X, Mei C-L, Chen D-G, Li J-H (2016) Feature selection in mixed data: A method using a novel fuzzy rough set-based information entropy. Pattern Recognit 56:1–157. https://doi.org/10.1016/j.patcog.2016.02.013
46. Yang Y-Y, Chen D-G, Wang H, Wang X-Z (2018) Incremental perspective for feature selection based on fuzzy rough sets. IEEE Trans Fuzzy Syst 26(3):1257–1273. https://doi.org/10.1109/TFUZZ.2017.2718492
47. Ji W-T, Pang Y, Jia X-Y, Wang Z-W, Hou F, Song B-Y, Liu M-Z, Wang R-L (2021) Fuzzy rough sets and fuzzy rough neural networks for feature selection: a review. Wires Data Min Knowl 11(3):e1402. https://doi.org/10.1002/widm.1402
48. Li L-Q, Wang X-L, Liu Z-X, Xie W-X (2019) A novel intuitionistic fuzzy clustering algorithm based on feature selection for multiple object tracking. Int J Fuzzy Syst 21(5):1613–1628. https://doi.org/10.1007/s40815-019-00645-7
49. Wan J-H, Chen H-M, Li T-R, Yuan Z, Liu J, Huang W (2021) Interactive and complementary feature selection via fuzzy multigranularity uncertainty measures. IEEE Trans Cybern. https://doi.org/10.1109/TCYB.2021.3112203
50. Ni P, Zhao S-Y, Wang X-Z, Chen H, Li C-P, Tsang ECC (2020) Incremental feature selection based on fuzzy rough sets. Inf Sci 536:185–204. https://doi.org/10.1016/j.ins.2020.04.038
51. Sun L, Wang T-X, Ding W-P, Xu J-C, Lin Y-J (2021) Feature selection using Fisher score and multilabel neighborhood rough sets for multilabel classification. Inf Sci 578:887–912. https://doi.org/10.1016/j.ins.2021.08.032
52. Qian W-B, Xiong C-Z, Wang Y-L (2021) A ranking-based feature selection for multi-label classification with fuzzy relative discernibility. Appl Soft Comput 102:106995. https://doi.org/10.1016/j.asoc.2020.106995
53. An S, Zhao E-H, Wang C-Z, Guo G, Zhao S-Y, Li P-Y (2021) Relative fuzzy rough approximations for feature selection and classification. IEEE Trans Cybern. https://doi.org/10.1109/TCYB.2021.3112674
54. Wang C-Z, Hu Q-H, Wang X-Z, Chen D-G, Qian Y-H, Dong Z (2018) Feature selection based on neighborhood discrimination index. IEEE Trans on Netw 29(7):2986–2999. https://doi.org/10.1109/TNNLS.2017.2710422
55. Dogan O, Kem FC, Oztaysi B (2022) Fuzzy association rule mining approach to identify e-commerce product association considering sales amount. Complex Intell Syst. https://doi.org/10.1007/s40747-021-00607-3
56. Dalkilic O (2021) On topological structures of virtual fuzzy parametrized fuzzy soft sets. Complex Intell Syst. https://doi.org/10.1007/s40747-021-00378-x
57. Wang C-Z, Shao M-W, He Q, Qian Y-H, Qi Y-L (2016) Feature subset selection based on fuzzy neighborhood rough sets. Knowl Based Syst 111(1):173–179. https://doi.org/10.1016/j.knosys.2016.08.009
58. Shreevastava S, Tiwari AK, Som T (2018) Intuitionistic fuzzy neighborhood rough set model for feature selection. Int J Fuzzy Syst Appl 7(2):75–84. https://doi.org/10.4018/IJFSA.2018040104
59. Wang C-Z, Huang Y, Shao M-W, Fan X-D (2019) Fuzzy rough set-based attribute reduction using distance measures. Knowl Based Syst 164:205–212. https://doi.org/10.1016/j.knosys.2018.10.038
60. Sang B-B, Chen H-M, Yang L, Li T-R, Xu W-H (2021) Incremental feature selection using a conditional entropy based on fuzzy dominance neighborhood rough sets. IEEE Trans Fuzzy Syst. https://doi.org/10.1109/TFUZZ.2021.3064686
61. Xu J-C, Wang Y, Mu H-Y, Huang F-Z (2018) Feature genes selection based on fuzzy neighborhood conditional entropy. J Intell Fuzzy Syst 36(1):117–126. https://doi.org/10.3233/JIFS-18100
62. Zhang X, Mei C-L, Chen D-G, Yang Y-Y, Li J-H (2019) Active incremental feature selection using a fuzzy-rough-set-based information entropy. IEEE Trans Fuzzy Syst 28(5):901–915. https://doi.org/10.1109/TFUZZ.2019.2959995
63. Sun L, Wang L-Y, Ding W-P, Qian Y-H, Xu J-C (2021) Feature selection using fuzzy neighborhood entropy-based uncertainty measures for fuzzy neighborhood multigranulation rough sets. IEEE Trans Fuzzy Syst 29(1):19–33. https://doi.org/10.1109/TFUZZ.2020.2989098
64. Xu J-C, Yuan M, Ma Y-Y (2021) Feature selection using self-information and entropy-based uncertainty measure for fuzzy neighborhood rough set. Complex Intell Syst. https://doi.org/10.1007/s40747-021-00356-3
65. Xu J-C, Shen K-L, Sun L (2022) Multi-label feature selection based on fuzzy neighborhood rough sets. Complex Intell Syst. https://doi.org/10.1007/s40747-021-00636-y
66. Xu J-C, Wang Y, Xu K-Q, Zhang T-L (2019) Feature genes selection using fuzzy rough uncertainty metric for tumor diagnosis. Comput Math Methods Med. https://doi.org/10.1155/2019/6705648
67. Xu J-C, Wang Y, Mu H-Y, Huang F-Z (2019) Feature genes selection based on fuzzy neighborhood conditional entropy. J Intell Fuzzy Syst 36(1):117–126. https://doi.org/10.3233/JIFS-18100
68. Liu X-M, Shen C, Wang W, Guan X-H (2020) CoEvil: a coevolutionary model for crime inference based on fuzzy rough feature selection. IEEE Trans Fuzzy Syst 28(5):806–817. https://doi.org/10.1109/TFUZZ.2019.2939957
69. Carvajal O, Melin P, Miramontes I, Prado-Arechiga G (2021) Optimal design of a general type-2 fuzzy classifier for the pulse level and its hardware implementation. Eng Appl Artif Intell 97. https://doi.org/10.1016/j.engappai.2020.104069
70. Pozna C, Precup RE (2014) Applications of signatures to expert systems modeling. Acta Polytech Hung 11(2):21–39
71. Borlea ID, Precup RE, Borlea AB, Iercan D (2021) A unified form of fuzzy C-means and K-means algorithms and its partitional implementation. Knowl Based Syst 214:106731. https://doi.org/10.1016/j.knosys.2020.106731
72. Garg H, Atef M (2022) Cq-ROFRS: covering q-rung orthopair fuzzy rough sets ant its application to multi-attribute decision-

making process. Complex Intell Syst. https://doi.org/10.1007/s40747-021-00622-4
73. Hu M, Tsang ECC, Guo Y-T, Chen D-G, Xu W-H (2022) Attribute reduction based on overlap degree and knearest neighbor rough sets in decision information systems. Knowl Based Syst 584:301–324. https://doi.org/10.1016/j.ins.2021.10.063
74. Hu M, Tsang ECC, Guo Y-T, Xu W-H (2021) Fast and robust attribute reduction based on the separability in fuzzy decision systems. IEEE Trans Cybern. https://doi.org/10.1109/TCYB.2020.3040803
75. Simon N, Friedman J, Hastie T, Tibshirani R (2013) A sparse-group lasso. J Comput Graph Stat 22(2):231–245. https://doi.org/10.1080/10618600.2012.681250
76. Demsar J, Schuurmans D (2006) Statistical comparison of classifiers over multiple data sets. J Mach Learn Res 7:1–30