Applied Intelligence (2025) 55:268
https://doi.org/10.1007/s10489-024-06134-1

Granular-ball-matrix-based incremental semi-supervised feature selection approach to high-dimensional variation using neighbourhood discernibility degree for ordered partially labelled dataset

Weihua Xu[1] · **Jinlong Li**[1]

Accepted: 29 November 2024

Abstract
In numerous real-world applications, data tends to be ordered and partially labelled, predominantly due to the constraints of labeling costs. The current methodologies for managing such data are inadequate, especially when confronted with the challenge of high-dimensional datasets, which often require reprocessing from the start, resulting in significant inefficiencies. To tackle this, we introduce an incremental semi-supervised feature selection algorithm that is grounded in neighborhood discernibility, and incorporates pseudolabel granular balls and matrix updating techniques. This novel approach evaluates the significance of features for both labelled and unlabelled data independently, using the power of neighborhood distinguishability to identify the most optimal subset of features. In a bid to enhance computational efficiency, especially with large datasets, we adopt a pseudolabel granular balls technique, which effectively segments the dataset into more manageable samples prior to feature selection. For high-dimensional data, we employ matrices to store neighborhood information, with distance functions and matrix structures that are tailored for both low and high-dimensional contexts. Furthermore, we present an innovative matrix updating method designed to accommodate fluctuations in the number of features. Our experimental results conducted across 12 datasets-including 4 with over 2000 features-demonstrate that our algorithm not only outperforms existing methods in handling large samples and high-dimensional datasets but also achieves an average time reduction of over six fold compared to similar semi-supervised algorithms. Moreover, we observe an average improvement in accuracy of 1.4%, 0.6%, and 0.2% per dataset for SVM, KNN, and Random Forest classifiers, respectively, when compared to the best-performing algorithm among the compared algorithms.

Keywords Granular ball · Incremental algorithm · Neighbourhood discernibility degree · Matrix update methodx · Ordered partially labelled data set

1 Introduction

In the modern, data-driven age, the surge of large-scale and high-dimensional data sets is a direct result of enhancements in data collection technologies. Feature selection [1, 2], the process of extracting a subset of the most effective features from a data set, has become indispensable in data mining for managing vast volumes of data. Particularly, we encounter dynamic data sets where the number of features grows over time, necessitating specialized feature selection techniques. The incremental feature reduction algorithm [3, 4] is a prevalent method for such evolving data sets. However, real-world data, such as medical diagnosis records, often come in ordered sequences with partial labeling due to the high costs of comprehensive labeling. For example, medical indicators like blood pressure and blood sugar levels remain unlabelled until a medical professional interprets them within the context of a diagnosis. This ordered and partially labelled nature is typical for many data types. The dimensions of a patient's sample expand when a doctor identifies a complex

✉ Weihua Xu
chxuwh@gmail.com

Jinlong Li
ljl010922@email.swu.edu.cn

[1] College of Artificial Intelligence, Southwest University, Chongqing 400715, Chongqing, China

Published online: 04 January 2025

condition, necessitating further examination of the lesion using sophisticated instruments. Unfortunately, the labeling of such detailed patient data is often limited, with only a small portion being labelled due to the associated expenses and labor. Thus, the analysis of partially labelled, dynamic, ordered data is not only inevitable but also critical in practical scenarios. Despite this, research on feature selection for these complex data types remains underdeveloped. Current studies have largely focused on either ordered or partially labeled data in isolation, failing to tackle their combined challenges. Furthermore, existing algorithms often struggle with the computational demands of large-scale, high-dimensional data. In light of these challenges, our team has initiated a study aimed at developing feature selection algorithms that are both efficient and effective for the complex and nuanced nature of partially labeled, dynamic, ordered data.

Rough set theory, a prominent data mining technique, is effective in partitioning datasets to reveal underlying relationships and patterns among attributes. It has been widely adopted in data mining [5], pattern recognition [6, 7], and knowledge discovery [8, 9], particularly excelling in scenarios involving incomplete, uncertain, or fuzzy data. In the context of partially labelled datasets with numerous missing labels, rough set theory proves highly effective in eliminating redundant features and assessing attribute significance. Among various models, the neighborhood rough set model is particularly noteworthy for data analysis due to its ability to operate without requiring any prior knowledge. Neighbourhood rough set [10, 11] is more robust in the face of noisy data and outliers as it measures the similarity or difference between attributes through the neighbourhood space. This capability allows the model to more effectively preserve valuable features while disregarding noisy ones during the feature selection process [12, 13]. Therefore, this paper introduces the neighborhood rough set model as a powerful tool for data analysis.

In addition, when tackling feature selection for partially labelled data, semi-supervised feature selection algorithms undoubtedly serve as powerful tools. With the growing prevalence of partially labelled data, more scholars are delving into research on semi-supervised feature selection. For example, Liu et al. [14] explored an attribute reduction based on neighborhood rough set over partially labelled data. Shu et al. [15] presented a neighbourhood discernibility degree-based semi-supervised feature selection algorithm and applies the granular ball technique can be used to process mixed types of partially labelled data. Huang et al. [16] proposed a semi-supervised attribute reduction for partially labelled categorical data which predict unlabelled data by means of the conditional probability. A supervised information granulation strategy is proposed by Liu et al. [17], which uses intra-class radius and extra-class radius to add or delete samples from the neighborhood. Gao et al. [18] presented a rough set-based semi-supervised attribute reduction method for partially labelled data. By studying the aforementioned literature, it is evident that the majority of semi-supervised feature selection algorithms are tailored for static mixed-type data, with limited attention given to ordered data. This gap highlights the urgent need for a feature selection algorithm that can effectively manage dynamically evolving, partially labelled ordered dataąła challenge that this paper aims to address as its primary contribution.

Dynamic feature selection algorithms are primarily employed to process streaming data or datasets that evolve over time. Their main objective is to adapt to changing data and environments by dynamically selecting relevant attributes throughout the training process. In recent years, there has been substantial research focused on dynamic feature selection, reflecting the growing importance of these algorithms in handling continuously changing data landscapes. Pan et al. [19] developed an incremental approach to feature selection using the weighted dominance-based neighborhood rough sets. Xu's team [20] described a matrix-based feature selection approach which can handle ordered data set with time-evolving features by conditional entropy. Yang et al. [21] built a feature selection framework based on discernibility score and proposed an incremental feature selection method using sample selection and feature-based accelerator. Yang's team [4] presented a novel incremental attribute reduction based on quantitative dominance-based neighborhood self-information for dynamic hybrid ordered decision system. Cai et al. [22] provided the incremental algorithms for updating reducts by making full use of the existing results from dynamic covering decision information systems. After collating and studying the aforementioned algorithms, we found that there are very few that address partially labelled ordered data within the context of dynamic feature selection. Therefore, this paper aims to bridge this gap by developing a dynamic feature selection algorithm specifically designed to handle partially labelled ordered datasets.

Feature evaluation criterion is a measure of the contribution or importance of a feature to the performance of a machine learning model. In rough set theory, these criteria can be divided into three main categories: consistency-based feature evaluation criteria [23, 24], information theory-based feature evaluation criteria [20, 25, 26] and discernibility viewpoint-based feature evaluation criteria [27, 28]. Among them, the discernibility degree, as an important concept of rough sets, is used to measure the ability to distinguish between two categories under specific conditions. This concept has gained significant attention and has been widely applied in recent years. Lin et al. [29] introduced a recursive method to gradually construct the multiple discernibility matrix by composing the refined discernibility matrix and

incremental discernibility matrix based on previous matrices. An attribute reduction with variable precision is proposed by Li et al. [30] which demonstrated the concepts of both indiscernibility and discernibility relations involving uncertain or imprecise information. By applying an information updating tool, Liu et al. [27] proposed the discernibility matrix based incremental feature selection algorithm, which greatly reduces the computation time. Inspired by the aforementioned article, this paper applies a dominant neighbourhood discernibility degree to a partially labelled ordered dataset, aiming to identify high-performance features within the dataset.

To improve the efficiency of feature selection, various techniques have been employed to mitigate the time complexity of the algorithm. Granular balls have captured our interest as a novel method to expedite feature selection. The concept of granular balls was initially introduced by Xia et al. They devised a sample segmentation method named granular ball, capable of completing the information granulation process without relying on any parameters [31, 32]. Additionally, Xia et al. [33] extended the granular ball model to neighborhood rough sets, enabling the generation of different neighborhoods for each object in an adaptive fashion. They further devised a corresponding feature selection algorithm based on this extension. From this point of view, Shu et al. [15] proposed a pseudolabelled granular ball generation algorithm to speed up the feature selection process. Chen et al. [34] constructed an granular ball guided selector attribute reduction algorithm by stepwise generation of granular balls. In this study, we introduce a pseudolabelled granular ball generation algorithm used to divide the overall sample into individual granular balls and use these granular balls to filter out a subset of features with the best performance in the form of integrated voting. Besides, Yang et al. [20] built a matrix - based dominant relation reduction method and Zhang et al. [35] used the Euclidean distance matrix to store and calculate the distance between samples in each neighbourhood. Drawing inspiration from their work, this paper introduces a matrix updating method that effectively leverages initial calculation results to accelerate the feature selection process.

This paper's primary contributions include:

(1) Introduction of a dominant neighborhood discernibility degree to evaluate the significance of candidate features for partially labelled ordered data.
(2) Development of the pseudolabelled granular ball technique and a matrix updating method effectively speed up the feature selection process compared to the four algorithms mentioned in the paper.
(3) Proposition of two feature selection algorithms tailored for partially labelled ordered data: a static algorithm NGM and a dynamic attribute algorithm with dynamically increasing features INGM-A.
(4) Presentation of extensive experimental results and statistical analyses demonstrating the superiority of the proposed algorithm over others in terms of computational time, classification accuracy, and feature subset size.

The remainder of this paper consists of the following sections. In Section 2, some basic concepts about neighbourhood rough sets and the definition of dominance relations are briefly introduced. A distance function to measure similarity and difference between samples is presented, along with the introduction of the related concept of a matrix to simplify calculations. In Section 3, introduce a dominant neighborhood discernibility degree to evaluate the significance of candidate features for partially labelled ordered data. In Section 4, a feature selection algorithm based on dominant neighbourhood discernibility degree via a pseudolabel granular ball selector and distance matrix updating methods is constructed for the ordered partially labelled decision system. In Section 5, a dynamic incremental feature selection algorithm for data dimensionality increase over time is established based on the theoretical foundation of the previous sections. In Section 6, a series of experiments on datasets such as those from UCI and Kaggle were conducted to demonstrate the effectiveness and feasibility of the proposed algorithm. The conclusions are presented in Section 7. The framework of the article is shown in Fig. 1.

2 Preliminaries

In this section, firstly, some fundamental concepts such as neighbourhoods, partially labelled ordered data, dominance relations etc., are briefly introduced. Relevant knowledge can be found in [8, 10, 20].

In rough set theory, data can usually be represented by an information system $IS = (U, A, V, f)$, where

(1) $U = \{x_1, x_2, ..., x_n\}$ denotes the set of objects and contains all the samples or examples;
(2) $A = \{a_1, a_2, ..., a_m\}$ denotes the set of features, which contains all the properties that describe an object, where a_j represents a single feature;
(3) $V = \cup_{a \in A} V_a$ is the collection of all feature values, where V_{a_j} is the set of possible values for feature a_j;
(4) $f : U \times A \longrightarrow V$ is an information function. $\forall x \in U, a \in A$, the relation function f gives the value of x for a_j, $f(x, a) \in V_a$.

More specifically, if the feature set in the aforementioned information system satisfies $A = C \cup D$ and $C \cap D = \varnothing$, the information system is a decision system $DS = (U, A = C \cup D, V, f)$, where $C = \{a_1, a_2, \ldots, a_m\}$

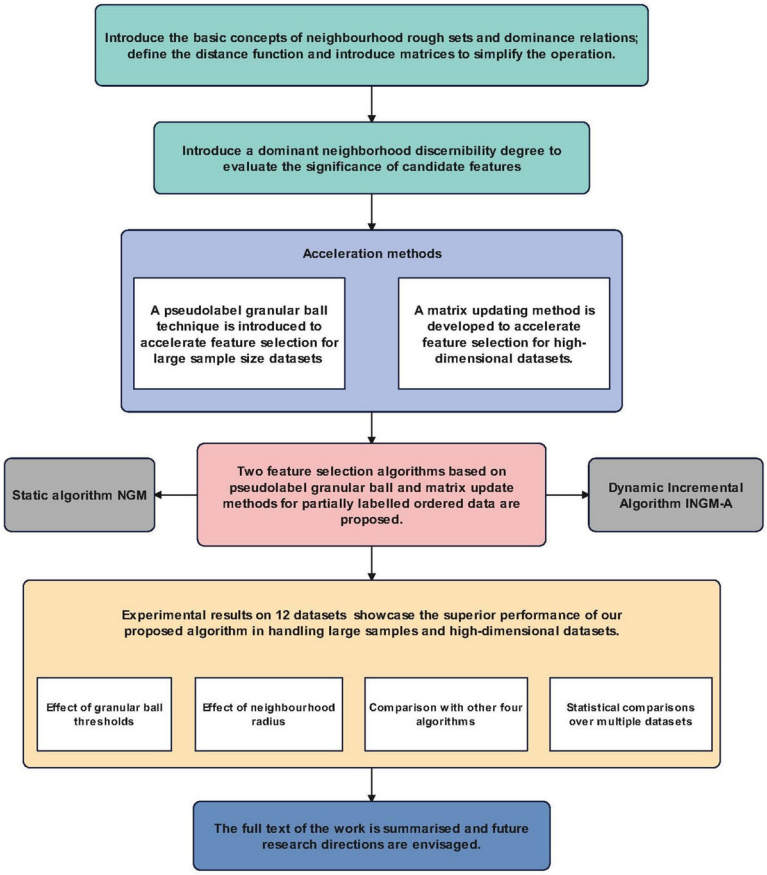


Fig. 1 Framework diagram of the article

represents the set of condition features, and $D = d$ denotes the decision feature. For $\forall B \subseteq C$, the feature subset determines an equivalence relation, which is denoted by $IND_B = \{(x, y) \in U \times U \mid \forall b \in B, f(x, b) = f(y, b)\}$. The equivalence relation IND_B partitions U into multiple equivalence classes, denoted by $U/IND(B) = \{X_1, X_2, \ldots, X_j\}$, where $X_i (1 \leq i \leq j)$ is the equivalence class.

If $U = U^l \cup U^u$, where U^u represents the collection of unlabelled objects, which means that for $\forall x \in U^u$, $f(x, d) = *$; U^l is the collection of decision label objects. Then the decision system is called the partially labelled decision system $LDS = (U = U^l \cup U^u, A = C \cup D, V, f)$.

In everyday life, many features are inherently ordered, and it is through this ordering that we can assess the quality of an item and categorize it. For instance, doctors evaluate a patient's health based on factors such as blood pressure, blood sugar, and pulse rate, all of which are ordered features. In a partially labelled decision system, when the domain of a feature is arranged in ascending or descending order and all features are criteria, an ordered partially labelled decision system ($OLDS$) is formed, denoted by $OLDS^{\geq} = \left(U = U^l \cup U^u, A = C \cup D, V, f\right)$.

Let $OLDS^{\geq} = \left(U = U^l \cup U^u, A = C \cup D, V, f\right)$ be an ordered partially labelled decision system, $\forall B \subseteq C, B \neq \emptyset$, the conditional relation with ascending order O_B (also called dominance relation) can be defined as

$$O_B = \{(x, y) \in U \times U : f(x, a) \geq f(y, a), \forall a \in B\}. \quad (1)$$

and the relational sets of x called B-dominated sets is defined as

$$O_B(x) = \{y \in U : x O_B y\}. \quad (2)$$

Example 1 In the following, an ordered partially labelled decision system on patient diagnosis is shown in Table 1. In this table, there are eight patients examined for breast cancer, i.e., $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$ denotes the collection of objects, where $U^u = \{x_2, x_3, x_4, x_6\}$ is a set of unlabelled objects, and $U^l = \{x_1, x_5, x_7, x_8\}$ is a set of labelled objects, where there are two classes, namely $Healthy$ and $Patient$. Since the condition feature set { BMI(unit: kg/m^2), Glucose(unit: mg/dL), Insulin(unit: μU/mL), HOMA, Leptin(unit: ng/mL) } has the numerical feature values, and the feature values in the parentheses are normalized into [0, 1],

i.e, $C = \{a_1, a_2, a_3, a_4, a_5\}$. Based on the partially labelled data U^u and U^l, let $B = C$, the dataset can be further divided by the dominance relation O_B to obtain the following results. $O_B^{U^l}(x_1) = \{x_1\}$; $O_B^{U^l}(x_5) = \{x_1, x_5, x_8\}$; $O_B^{U^l}(x_7) = \{x_1, x_7, x_8\}$; $O_B^{U^l}(x_8) = \{x_8\}$; $O_B^{U^u}(x_2) = \{x_2, x_3, x_4, x_6\}$; $O_B^{U^u}(x_3) = \{x_3\}$; $O_B^{U^u}(x_4) = \{x_4, x_6\}$; $O_B^{U^u}(x_6) = \{x_6\}$.

Given an ordered partially labelled decision system $OLDS^{\geq} = \left(U = U^l \cup U^u, A = C \cup D, V, f\right)$, for $\forall B \subseteq C$, the neighbourhood relation NR_B can be defined as

$$NR_B = \{(x, y) \in U \times U \mid \triangle_B(x, y) \leq \delta\}, \tag{3}$$

where $\triangle_B(x, y)$ denotes a distance function that measures the similarity of two objects. δ called the neighbourhood radius.

NR_B has the following properties, when it is a neighbourhood relation in an information systems.

(1) Reflexive: $\forall x \in U$, then xNR_Bx;
(2) Symmetric: $\forall x, y \in U$, let xNR_By, then yNR_Bx;
(3) Non-transitive: $\forall x, y, z \in U$, let xNR_By and yNR_Bz, then xNR_Bz cannot be taken as true;

The neighbourhood relation induces a set of coverage of the universe U. For $U = \{x_1, x_2, \ldots, x_n\}$, $B \subseteq C$, the classification induced by the universe U under the neighbourhood relation NR_B is $U/NR_B = \{\partial_B(x_1), \partial_B(x_2), \ldots, \partial_B(x_n)\}$, where $\partial_B(x) = \{y \in U \mid (x, y) \in NR_B\}$ denotes the neighbours of each object within a given range on feature set B. For simplicity, U/NR_B will be replaced by U/B.

In data analytics mining and machine learning, it's crucial to understand the degree of dissimilarity among various data entities and subsequently assess their distance and categorization. Drawing inspiration from Hu et al. [36], we introduced a method to quantify the distance between objects in an ordered partially labelled decision system, which is defined as follows.

Let $OLDS^{\geq} = \left(U = U^l \cup U^u, A = C \cup D, V, f\right)$ be an ordered partially labelled decision system. For each $a \in C$ and $x, y \in U$ the distance between x and y w.r.t. the feature a is defined as

$$d_a(x, y) = \frac{f(x, a) - f(y, a)}{\max_a - \min_a}, \quad f(x, a) \geq f(y, a), \tag{4}$$

where $\max_a$ and $\min_a$ are the maximum and minimum values of feature a, respectively.

It is widely acknowledged that when two objects are closer in proximity, they are more likely to be similar and belong to the same object class. In [37], Yuan et al. introduced a heterogeneous Euclidean-overlap metric to handle numerical data. Building upon the heterogeneous Euclidean-overlap metric method, the computational framework for calculating the distance between homogeneous data is constructed as follows.

Definition 1 Let $OLDS^{\geq} = \left(U = U^l \cup U^u, A = C \cup D, V, f\right), C = \{a_1, a_2, ..., a_l\}$ be an ordered partially labelled decision system where l is the total number of features in C. For each $x, y \in U$ and $a_{l'} \in B \subseteq C\left(l' \in \{1, 2, .., l\}\right)$ the Euclidean distance metric (EDM) on feature $a_{l'}$ is defined as

$$EDM_B(x, y) = \begin{cases} \begin{cases} \sqrt{\frac{1}{l}\sum_{l'=1}^{l'} d_{a_{l'}}^2(x, y)}, & y \in O_B(x) \\ 1, & y \notin O_B(x) \end{cases}, & x \neq y \\ 0, & x = y \end{cases} \tag{5}$$

where l' is the number of features and $|B| = l'$, l is the total number of features and $|C| = l$, $|*|$ represents the cardinality of a set, $d_{a_{l'}}(x, y)$ is the distance between objects x and y under the feature $a_{l'}$.

Based on Definition 1, $(EDM_B)_M = (EDM_B(x, y))_{n \times n}$ is called Euclidean distance metric matrix of feature a can be easily calculated.

Example 2 (Continued from Example 1). For the ordered partially labelled decision system in Table 1, let $B = C$, the distances between x_1 and x_2 w.r.t. $a_i (i = 1, 2, 3, 4, 5)$ are computed as follows. $d_{a_1}(x_2, x_1) = 0.42$, $d_{a_2}(x_2, x_1) = 0.98$, $d_{a_3}(x_2, x_1) = 1$, $d_{a_4}(x_2, x_1) = 1$, $d_{a_5}(x_2, x_1) = 0.97$,

Table 1 An ordered partially labelled decision system for patient diagnosis (normalized results)

Sample	BMI a_1	Glucose a_2	Insulin a_3	HOMA a_4	Leptin a_5	Class d
x_1	23.5 (0.22)	70 (0.00)	2.707 (0.00)	0.467 (0.00)	8.807 (0.03)	Healthy
x_2	27.7 (0.64)	196 (0.98)	51.814 (1.00)	25.050 (1.00)	70.882 (1.00)	*
x_3	23.1 (0.18)	91 (0.16)	4.498 (0.04)	1.010 (0.02)	17.939 (0.17)	*
x_4	23.0 (0.17)	83 (0.10)	4.952 (0.05)	1.014 (0.02)	17.127 (0.16)	*
x_5	31.3 (1.00)	199 (1.00)	12.162 (0.19)	5.970 (0.22)	18.131 (0.17)	Patient
x_6	22.7 (0.13)	77 (0.05)	4.690 (0.04)	0.891 (0.02)	6.964 (0.00)	*
x_7	29.7 (0.84)	85 (0.12)	14.649 (0.24)	3.071 (0.11)	26.517 (0.31)	Patient
x_8	21.4 (0.00)	77 (0.05)	3.226 (0.01)	0.613 (0.01)	9.883 (0.05)	Healthy

and so on. Then, the Euclidean distance metric matrix of the decision system can be calculated as follows.

$$(EDM_B)_M = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0.90 & 0 & 0.83 & 0.84 & 1 & 0.89 & 1 & 0.91 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0.11 \\ 1 & 1 & 1 & 0 & 1 & 0.08 & 1 & 0.09 \\ 0.59 & 1 & 0.54 & 0.56 & 0 & 0.59 & 1 & 0.63 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0.33 & 1 & 1 & 0.32 & 1 & 0.36 & 0 & 0.41 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

In the above, we give a distance function to measure the similarity and difference between two objects. According to (3), we know that when the distance between two objects is less than the radius of the neighbourhood δ, these two objects satisfy the neighbourhood relationship and can be classified into the same class. In other words, we have represented the similarity between these eight patients with a distance matrix through the B feature set. Given a neighbourhood radius δ, it is possible to know what are the other patients with similar conditions to the patient.

By Definition 1, the value of the EDM is bounded between 0 and 1, so the value domain of the neighbourhood radius δ should also belong to [0, 1]. Therefore, in the ordered partially labelled decision system, the new neighbourhood relationship can be defined as follows.

Definition 2 Given an ordered partially labelled decision system $OLDS^{\geq} = \left(U = U^l \cup U^u, A = C \cup D, V, f\right)$, let $\forall B \subseteq C, \delta \in [0, 1]$, and for $\forall x, y \in U$, a neighbourhood relation can be defined by

$$N_B^{\delta} = \{(x, y) \in U \times U : EDM_B(x, y) \leq \delta\}. \tag{6}$$

Thus the neighbourhood relation N_B^{δ} of U can be represented by a relation matrix $(NR_B)_M = (NR_B(x, y))_{n \times n}$, where

$$(NR_B)(x, y) = \begin{cases} 1, & EDM_B(x, y) \leqslant \delta, \\ 0, & \text{else.} \end{cases} \tag{7}$$

Proposition 1 Let $OLDS^{\geq} = (U = U^l \cup U^u, A = C \cup D, V, f)$ be an ordered partially labelled decision system. Given $B_1, B_2 \subseteq C, \delta_1, \delta_2 \in [0, 1]$, neighbourhood relation N_B^{δ},the system then has the following properties:

(1) If $B_1 \subseteq B_2$, then $N_{B_2}^{\delta}(x) \subseteq N_{B_1}^{\delta}(x)$.
(2) If $\delta_1 \leq \delta_2$, then $N_B^{\delta_1}(x) \subseteq N_B^{\delta_2}(x)$.

The above two properties are fundamental in neighbourhoods and the proof is obvious.

Example 3 (Continued from Example 2). According to Definition 1 and 2, let $B = \{a_1, a_2, a_3, a_4, a_5\}$, in order to make the example more concise and without too many neighbourhood elements in each sample, set $\delta = 0.35$, then

$$(EDM_B)_M = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0.90 & 0 & 0.83 & 0.84 & 1 & 0.89 & 1 & 0.91 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0.11 \\ 1 & 1 & 1 & 0 & 1 & 0.08 & 1 & 0.09 \\ 0.59 & 1 & 0.54 & 0.56 & 0 & 0.59 & 1 & 0.63 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0.33 & 1 & 1 & 0.32 & 1 & 0.36 & 0 & 0.41 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

$$\Longrightarrow (NR_B)_M = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

According to the above matrix, we can get $N_B^{\delta}(x_1) = \{x_1\}$; $N_B^{\delta}(x_2) = \{x_2\}$; $N_B^{\delta}(x_3) = \{x_3, x_8\}$; $N_B^{\delta}(x_4) = \{x_4, x_6, x_8\}$; $N_B^{\delta}(x_5) = \{x_5\}$; $N_B^{\delta}(x_6) = \{x_6\}$; $N_B^{\delta}(x_7) = \{x_1, x_4, x_7\}$; $N_B^{\delta}(x_8) = \{x_8\}$.

Therefore, $U/N_B^{\delta} = \{\{x_1\}, \{x_2\}, \{x_3, x_8\}, \{x_4, x_6, x_8\}, \{x_5\}, \{x_6\}, \{x_1, x_4, x_7\}, \{x_8\}\}$, i.e. given a neighbourhood radius $\delta = 0.35$, eight patients can be classified into the above groups. The reason $(EDM_B)_M$ is not a symmetric matrix is that the matrix is defined under the dominance relation $O_B(x)$. In this case, patients will prefer to be closer and more similar to those with higher values of the five indicators, and more distant from those with lower values of the indicators.

The following is a brief introduction to the concept of granulation ball. The concept was initially introduced by Xia et al. They devised a sample segmentation method named granular ball, capable of completing the information granulation process without relying on any parameters [31, 32]. The relevant definitions of granulation ball are as follows.

Definition 3 [33] (Granular ball) Given a decision system $DS = (U, A = C \cup D, V, f), \forall B \subseteq C$, let $GB_B^U = \{GB_1, GB_2, \ldots, GB_n\}$ be the collection of all granular balls induced by B under $U, \forall GB_i \in GB_B^U (1 \leq i \leq n)$ is represented by centre c and radius r. The centre c and radius r can be defined as

$$c = \frac{1}{|GB_i|} \sum_{j=1}^{|GB_i|} x_j, \tag{8}$$

and

$$r = \frac{1}{|GB_i|} \sum_{j=1}^{|GB_i|} \triangle \left(x_j, c\right), \tag{9}$$

where c denotes the centre of gravity of all objects in GB_i, and r is an average of the distances from all objects to c in GB_i. $\triangle \left(x_j, c\right)$ is the distance between object x_j in GB_i and the centre c under the feature subset B.

Furthermore, the purity of a granular ball is used to indicate the percentage of the maximum number of objects with the same label in the granular ball.

Definition 4 [31] (Purity) Given a decision system $DS = (U, A = C \cup D, V, f), \forall B \subseteq C$, a granular ball $GB_i \in GB_B^U$ and $GB_i/D = \{D_1, D_2, \ldots, D_m\}$, the purity P of GB_i can be defined as

$$P = \frac{\max_{j=1}^{m} \left|D_j\right|}{|GB_i|} \tag{10}$$

3 Neighbourhood discernibility degree for an OLDS

In this section, we will introduce the concept of neighbourhood discernibility degree to aid in measuring the importance of candidate features. Sheng et al. [28] presented the concept of neighbourhood discernibility degree by generalizing the traditional discernibility degree under the information system, and this paper extends it to ordered partially labelled data.

Definition 5 (Neighborhood discernibility degree) Given an ordered partially labelled decision system $OLDS^{\geq} = \left(U = U^l \cup U^u, A = C \cup D, V, f\right), \forall B \subseteq C$, neighbourhood relation N_B^δ and $U^u/N_B^\delta = \left\{N_B^\delta\left(x_1\right), N_B^\delta\left(x_2\right), \ldots N_B^\delta\left(x_m\right)\right\}, m = |U^u|$, the neighbourhood discernibility degree of feature subset B under the universe U^u is defined as

$$NDD^{U^u}(B) = \left|U^u\right|^2 - \sum_{i=1}^{m} \left|N_B^\delta\left(x_i\right)\right| \tag{11}$$

where $|*|$ represents the cardinality of a set.

The neighbourhood discernibility degree is a measure of the ability of a subset of features B to distinguish between the entire unlabelled data set. The larger value of $NDD^{U^u}(B)$ indicates that there are fewer elements in the neighbourhood of each unlabelled sample obtained through feature subset B, i.e., feature subset B is more capable of dividing the unlabelled sample.

Based on the neighbourhood discernibility degree, we can further obtain the definition of the relative neighbourhood discernibility degree.

Definition 6 (Relative neighborhood discernibility degree) Let $OLDS^{\geq} = \left(U = U^l \cup U^u, A = C \cup D, V, f\right)$ be a partially labelled decision system, $\forall B \subseteq C$, neighbourhood relation N_B^δ and $U^l/N_B^\delta = \left\{N_B^\delta\left(x_1\right), N_B^\delta\left(x_2\right), \ldots N_B^\delta\left(x_T\right)\right\}$, $T = \left|U^l\right|, \forall x \subseteq U^l$, $[x]_D$ represents the decision class of object x under decision feature D, the relative neighbourhood discernibility degree can be defined as

$$NDD^{U^l}(B, D) = \sum_{i=1}^{T} \left|N_B^\delta\left(x_i\right)\right| - \sum_{i=1}^{T} \left|N_B^\delta\left(x_i\right) \cap [x_i]_D\right| \tag{12}$$

The relative neighbourhood discernibility degree is used to denote the difference in the ability to classify samples by feature subset B and decision attribute D. A smaller value of $NDD^{U^l}(B, D)$ denotes a smaller difference in the samples classified by feature subset B and decision attribute D, respectively, i.e., the more proximity between the roles of feature subset B and decision attribute D.

Proposition 2 Given an ordered partially labelled decision system $OLDS^{\geq} = \left(U = U^l \cup U^u, A = C \cup D, V, f\right)$, $\forall B \subseteq C$, neighbourhood relation N_B^δ, $T = \left|U^l\right|, \forall x \subseteq U^l$, $[x]_D$ represents the decision class of object x under decision feature D, then $NDD^{U^u}(B)$ and $NDD^{U^l}(B, D)$ have the following properties:

(1) $0 \leq NDD^{U^u}(B) \leq |U^u|\left(|U^u| - 1\right)$;
(2) $0 \leq NDD^{U^l}(B, D) \leq \left|U^l\right|^2 - \sum_{i=1}^{T} |[x_i]_D|$;

Proof Let $U^u/N_B^\delta = \{\{x_1\}, \{x_2\}, \ldots, \{x_m\}\}$, the neighborhood discernibility degree of U^u will reach its maximum value $NDD^{U^u}(B) = |U^u|\left(|U^u| - 1\right)$. Let $U^u/N_B^\delta = \{U^u, U^u, \ldots, U^u\}$, the neighborhood discernibility degree of U^u will reach its minimum value $NDD^{U^u}(B) = 0$. Therefore, $0 \leq NDD^{U^u}(B) \leq |U^u|\left(|U^u| - 1\right)$. Similarly, we can get $0 \leq NDD^{U^l}(B) \leq \left|U^l\right|^2 - \sum_{i=1}^{T} |[x_i]_D|$.

Examples of calculating $NDD^{U^u}(B)$ and $NDD^{U^l}(B, D)$ are as follows.

Example 4 (Continued from Example 1). For the ordered partially labelled decision system in Table 1, $U^u = \{x_2, x_3, x_4, x_6\}, U^l = \{x_1, x_5, x_7, x_8\}$, let $B = \{a_1, a_2, a_3, a_4\}$. The sample size was reduced compared to Example 3 because the samples were divided into labelled and unlabelled samples

for separate calculations. To prevent a sample from having only itself in its neighbourhood, set $\delta = 0.85$, then

$$(EDM_B)_M^{U^u} = \begin{bmatrix} 0 & 0.85 & 0.87 & 0.89 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0.04 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

$$\Longrightarrow (NR_B)_M^{U^u} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$(EDM_B)_M^{U^l} = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0.81 & 0 & 1 & 0.82 \\ 0.57 & 1 & 0 & 0.60 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

$$\Longrightarrow (NR_B)_M^{U^l} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Thus,

$$\begin{aligned} NDD^{U^u}(B) &= \left|U^u\right|^2 - \sum_{i=1}^{m} \left|N_B^\delta(x_i)\right| \\ &= 4^2 - (2+1+2+1) \\ &= 10 \end{aligned}$$

$$\begin{aligned} NDD^{U^l}(B,D) &= \sum_{i=1}^{T} \left|N_B^\delta(x_i)\right| - \sum_{i=1}^{T} \left|N_B^\delta(x_i) \cap [x_i]_D\right| \\ &= (1+3+3+1) - (1+1+1+1) \\ &= 4 \end{aligned}$$

Example 4 demonstrates the ability to classify eight patients by a subset of features B. $NDD^{U^u}(B)$ represents the ability to classify four unknown categories of patients. The larger the value of $NDD^{U^u}(B)$, the better the ability of B to differentiate between these four patients. $NDD^{U^l}(B)$ represents the ability to classify four known categories of patients. The smaller the value of $NDD^{U^l}(B)$, the better the ability of B to correctly determine whether or not the four patients are sick.

The monotonicity of the neighborhood discernibility degree for selecting candidate features is discussed next, which is a crucial aspect guiding the feature selection search process.

Proposition 3 (Monotonicity) Given an ordered partially labelled decision system $OLDS^{\geq} = \left(U = U^l \cup U^u, A = C \cup D, V, f\right)$, for $\forall B_1, B_2 \subseteq C$, if $B_1 \subseteq B_2$, and the same neighbourhood radius δ, then $NDD^{U^u}(B_1) \leq NDD^{U^u}(B_2)$, $NDD^{U^l}(B_1, D) \geq NDD^{U^l}(B_2, D)$.

Proof $\forall x \in U$, since $B_1 \subseteq B_2$, by Proposition 1, we can easily obtain that $N_{B_2}^\delta(x_i) \subseteq N_{B_1}^\delta(x_i)$. Suppose $U^u/N_{B_1}^\delta = \left\{N_{B_1}^\delta(x_1), N_{B_1}^\delta(x_2), \ldots, N_{B_1}^\delta(x_m)\right\}$ and $U^u/N_{B_2}^\delta = \left\{N_{B_2}^\delta(x_1), N_{B_2}^\delta(x_2), \ldots, N_{B_2}^\delta(x_m)\right\}$, we have $\sum_{i=1}^{m}\left|N_{B_2}^\delta(x_i)\right| \leq \sum_{i=1}^{m}\left|N_{B_1}^\delta(x_i)\right|$. Therefore, it is easy to get $NDD^{U^u}(B_1) \leq NDD^{U^u}(B_2)$. Similarly, suppose $U^l/N_{B_1}^\delta = \left\{N_{B_1}^\delta(x_1), N_{B_1}^\delta(x_2), \ldots, N_{B_1}^\delta(x_n)\right\}$, $U^l/N_{B_2}^\delta = \left\{N_{B_2}^\delta(x_1), N_{B_2}^\delta(x_2), \ldots, N_{B_2}^\delta(x_n)\right\}$ and the decision class of $x\left(\forall x \in U^l\right)$ under decision feature D is $[x]_D$. Thus, $NDD^{U^l}(B_1, D) - NDD^{U^l}(B_2, D) = \sum_{i=1}^{T}\left|N_{B_1}^\delta(x_i)\right| - \sum_{i=1}^{T}\left|N_{B_1}^\delta(x_i) \cap [x_i]_D\right| - \sum_{i=1}^{T}\left|N_{B_2}^\delta(x_i)\right| + \sum_{i=1}^{T}\left|N_{B_2}^\delta(x_i) \cap [x_i]_D\right| = \sum_{i=1}^{T}\left|N_{B_1}^\delta(x_i)\right| - \sum_{i=1}^{T}\left|N_{B_2}^\delta(x_i)\right| - \left(\sum_{i=1}^{T}\left|N_{B_1}^\delta(x_i) \cap [x_i]_D\right| - \sum_{i=1}^{T}\left|N_{B_2}^\delta(x_i) \cap [x_i]_D\right|\right)$. Since $N_{B_2}^\delta(x) \subseteq N_{B_1}^\delta(x)$, then satisfy $\left|N_{B_1}^\delta(x)\right| - \left|N_{B_2}^\delta(x)\right| = \left|N_{B_1}^\delta(x) - N_{B_2}^\delta(x)\right|$ and $\left|N_{B_1}^\delta(x) \cap [x]_D\right| - \left|N_{B_2}^\delta(x) \cap [x]_D\right| = \left|\left(N_{B_1}^\delta(x) - N_{B_2}^\delta(x)\right) \cap [x]_D\right|$, so $NDD^{U^l}(B_1, D) - NDD^{U^l}(B_2, D) = \sum_{i=1}^{T}\left(\left|N_{B_1}^\delta(x_i) - N_{B_2}^\delta(x_i)\right|\right) - \sum_{i=1}^{T}\left|\left(N_{B_1}^\delta(x_i) - N_{B_2}^\delta(x_i)\right) \cap [x_i]_D\right| \geq 0$. Hence, $NDD^{U^l}(B_1, D) \geq NDD^{U^l}(B_2, D)$.

With Proposition 3, it is easy to construct a metric for evaluating the importance of features in the feature selection process, which is defined as follows.

Definition 7 (Inner and outer significance) Given an ordered partially labelled decision system $OLDS^{\geq} = \left(U = U^l \cup U^u, A = C \cup D, V, f\right), \forall B \subseteq C$, the inner significance of feature b can be defined as

$$\begin{aligned} SIG_{in}^U(b, B, D) = {} & NDD^{U^l}(B - \{b\}, D) - NDD^{U^l}(B, D) \\ & + NDD^{U^u}(B) - NDD^{U^u}(B - \{b\}) \end{aligned} \tag{13}$$

The inner significance can delete some redundant features from the candidate features. When $SIG_{in}^U(b, B, D) = 0$, it means that feature b is a redundant feature.

And given an ordered partially labelled decision system $OLDS^{\geq} = \left(U = U^l \cup U^u, A = C \cup D, V, f\right), \forall B \subseteq C$ and $b \in C - B$, the same reasoning can be used to obtain the outer significance of feature b

$$\begin{aligned} SIG_{out}^U(b, B, D) = {} & NDD^{U^l}(B, D) - NDD^{U^l}(B \cup \{b\}, D) \\ & + NDD^{U^u}(B \cup \{b\}) - NDD^{U^u}(B) \end{aligned} \tag{14}$$

Similarly, the outer significance can sift out important features from the feature set in the universe U to be added to

the candidate feature set. If $\mathrm{SIG}_{\mathrm{out}}^{U}(b, B, D) > 0$, feature b is deemed non-redundant and should be included in the candidate features. With the above definitions in place, we can deduce the conditions required for completing feature selection as follows.

Proposition 4 Let $OLDS^{\geq} = (U = U^l \cup U^u, A = C \cup D, V, f)$ be an ordered partially labelled decision system, $\forall B \subseteq C$, the feature subset B selected by outer and inner significance to become a selected feature subset needs to satisfy the following conditions:

(1) $NDD^{U^l}(C, D) = NDD^{U^l}(B, D), NDD^{U^u}(C) = NDD^{U^u}(B)$;
(2) $\forall b \in B, NDD^{U^l}(B, D) < NDD^{U^l}(B - \{b\}, D), NDD^{U^u}(B - \{b\}) < NDD^{U^u}(B)$;

The first condition states that the selected subset of features B has the same discriminability as the entire set of conditional features C to infer and the entire set of decision features D.

The second condition ensures that the selected features subset B is minimal and no redundant features exist.

4 Feature selection for ordered partially labelled data via pseudolabel granular ball and distance matrix updating methods

4.1 Pseudolabel granular ball

In this subsection, we introduce a pseudolabel granular ball generation algorithm. When dealing with large sample datasets, granular balls can effectively partition the dataset, consolidate information, and simplify computations to some extent.

Since the generation of granular balls requires labelled data, pseudolabeling techniques are crucial to apply them on partially labelled ordered datasets. Pseudolabeling tags data with a pseudolabel based on the information available in the partially labelled dataset itself. The specific definitions are as follows.

Definition 8 Given a decision system $DS = (U, A = C \cup D, V, f)$, the ordered partially labelled decision system labelled with pseudolabels can be defined as

$$OLDS^{\geq PL} = \left(U = U^l \cup U^u, C, D, D^{PL}, f\right), \tag{15}$$

where D^{PL} is the pseudolabel decision feature. $\forall x \in U, f\left(x, D_C^{PL}\right)$ represents the pseudolabel for object x, which can be generated from the learning method based on feature subset C. This paper adopts a known k-means clustering method to obtain pseudolabels.

With the introduction of the pseudolabeling technology, the purity of the granular ball was redefined as follows.

Definition 9 (Pseudolabel purity) Let $OLDS^{\geq PL} = \left(U = U^l \cup U^u, C, D, D^{PL}, f\right)$ be a pseudolabel decision system, $\forall B \subseteq C$, a granular ball $GB_i \in GB_B^U$ and $GB_i / D_C^{PL} = \left\{D_1^{PL}, D_2^{PL}, \ldots, D_m^{PL}\right\}$, the purity P^{PL} of GB_i is defined as

$$P^{PL} = \frac{\max_{j=1}^{m} \left|D_j^{PL}\right|}{|GB_i|} \tag{16}$$

After the above definition, we can obtain a pseudolabel granular ball generation algorithm applied to partially labelled ordered datasets as detailed below.

Algorithm 1 Pseudolabel Granular Ball Generation Algorithm for Ordered Partially Labelled Data (PGA algorithm).

Input: An ordered partially labelled decision system $OLDS^{\geq} = (U = U^l \cup U^u, A = C \cup D, V, f)$;
Output: A pseudolabel granular ball set GB_C^U.

1: Let $GB_C^U = \emptyset$
2: Obtain the pseudolabel of the object by using k-means clustering and replace the original label in the data with the pseudolabel
3: Consider the entire universe U as an initial granular ball
4: Decompose each pseudolabel granular ball by using two-means clustering
5: Compute the purity of each pseudolabel granular ball by Definition 9
6: **if** the purity of each pseudolabel granular ball reaches a given threshold **then**
7: Put each pseudolabel granular ball into GB_C^U
8: **else**
9: Return to Step 4
10: **end if**

The time complexity of the PGA algorithm can be derived from the time complexity of k-means clustering, whose time complexity is $O(Nkt)$ [31], where N represents the size of objects in the data, k represents the size of clustering, and t represents the number of iterations. In the process of generating granular balls, the time complexity of generating granular balls is close to $O(|U|)$ as the k-means algorithm generally has a fast convergence rate and can be considered to be close to linear. Moreover, we need to calculate the distance between two objects whose time complexity is $O(K \times |U| \times |C|)$. So, after analysing it is known that the time complexity of PGA is $O(K \times |U| \times |C|)$.

The following example illustrates how to generate a collection of pseudolabel granule balls.

Example 5 (Continuation of Example 1). An ordered partially labelled decision system is shown in Table 1, where

$U = \{x_1, x_2, \ldots, x_8\}$, $B = \{a_1, a_2, a_3, a_4, a_5\}$. Assuming that the given threshold is 0.7, the sample data is classified into two classes using the 2-means clustering algorithm since the number of labelled species in the original data is 2, which are $\{x_2, x_3, x_4, x_6, x_7, x_8\}$ and $\{x_1, x_5\}$, respectively. Setting the pseudolabel of $\{x_2, x_3, x_4, x_6, x_7, x_8\}$ to 0 and the pseudolabel of $\{x_1, x_5\}$ to 1.$U = \{x_1, x_2, \ldots, x_8\}$ is regarded as an initial pseudolabel granular ball.

According to Step 3, obtain two pseudolabel granular balls $\{x_2, x_3, x_4, x_6, x_7, x_8\}$ and $\{x_1, x_5\}$.

According to Step 4, the purity of pseudolabel granular balls $\{x_2, x_3, x_4, x_6, x_7, x_8\}$ and $\{x_1, x_5\}$ is 1 .

According to Step 5, the purity of pseudolabel granular balls $\{x_2, x_3, x_4, x_6, x_7, x_8\}$ and $\{x_1, x_5\}$ reaches the given threshold, and a pseudolabel granular ball set $GB_C^U = \{\{x_2, x_3, x_4, x_6, x_7, x_8\}, \{x_1, x_5\}\}$ is obtained.

4.2 A distance matrix updating methods

A key point in the algorithm that saves time is how to save the result of a previous calculation and how to update it for subsequent calculations. This paper uses $(EDM_B)_M^U$ to save the results of previous calculations. First, we describe in detail how the $(EDM_B)_M^U$ is updated when features are added.

Proposition 5 Given an ordered partially labelled decision system $OLDS^{\geq} = \left(U = U^l \cup U^u, A = C \cup D, V, f\right)$, $\forall B \subseteq C$ and $a_{l'} \in (C - B)$ $\left(l' \in \{1, 2, \ldots, n\}\right)$, for each $x, y \in U$, $\left(EDM_{B\cup a_{l'}}\right)_M^U$ can be calculated as follows:

$$EDM_{B\cup a_{l'}}(x,y) = \begin{cases} \sqrt{EDM_B^2(x,y) + \frac{1}{l} \cdot d_{a_{l'}}^2(x,y)}, & y \in O_B(x) \\ \sqrt{EDM_B^2(x,y) + 1}, & y \notin O_B(x) \end{cases} \tag{17}$$

Proposition 5 gives how to update the $(EDM_B)_M^U$ matrix when features are added, and the equation reveals the relationship between the newly added features and the originally selected features. This forms the basis for subsequent feature incremental algorithms. Here is an example of how to update $(EDM_B)_M^U$.

Example 6 (Continuation of Example 4). Let $B = \{a_1, a_2, a_3, a_4\}$ and $a_{l'} = a_5$, $\delta = 0.85$,then

$$\left((EDM_B)_M^{U^u}\right)^2 = \begin{bmatrix} 0 & 0.7211 & 0.7512 & 0.7984 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0.0012 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

$$\left(EDM_{B\cup a_{l'}}\right)_M^{U^u} = \begin{bmatrix} 0 & 0.9265 & 0.9448 & 0.9992 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0.0794 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

$$\Longrightarrow \left(NR_{B\cup a_{l'}}\right)_M^{U^u} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Thus,

$$\begin{aligned} NDD^{U^u}(B \cup a_{l'}) &= \left|U^u\right|^2 - \sum_{i=1}^{m} \left|N_{B\cup a_{l'}}^{\delta}(x_i)\right| \\ &= 4^2 - (1 + 1 + 2 + 1) \\ &= 11 \end{aligned}$$

where the parts highlighted in red represent the changes updated with the addition of the a_5 feature. As can be seen from the example, with the addition of the a_5 feature, patients x_1 and x_2 no longer belong to the same class with the same neighbourhood radius. It means that the feature subset B is more capable of distinguishing between the samples.

Similar to the Proposition 5, the updating principle of $(EDM_B)_M^U$ in the case of reducing features is as follows.

Proposition 6 Let $OLDS^{\geq} = \left(U = U^l \cup U^u, A = C \cup D, V, f\right)$ be an ordered partially labelled decision system, $\forall B \subseteq C$ and $a_{l'} \in B$, for each $x, y \in U$, $\left(EDM_{B-\{a_{l'}\}}\right)_M^U$ can be calculated as follows:

$$EDM_{B-\{a_{l'}\}}(x,y) = \begin{cases} \sqrt{EDM_B^2(x,y) - \frac{1}{l} \cdot d_{a_{l'}}^2(x,y)}, & y \in O_B(x) \\ \sqrt{EDM_B^2(x,y) - 1}, & y \notin O_B(x) \end{cases} \tag{18}$$

Proposition 6 clearly shows how the $(EDM_B)_M^U$ matrix is updated when redundant features need to be removed. The above two propositions can help us to reduce many repetitive computations, particularly beneficial when dealing with a large number of features in the dataset.

Even so, we still encounter challenges with high-dimensional ordered data. As the number of features in the sample data increases, there will be fewer and fewer samples satisfying the dominance relation $O_B(x)$. Consequently, a significant amount of valid information may be lost. Therefore, when confronted with high-dimensional samples, we adjust the definition of the distance function $d_a(x, y)$ and matrix computation $EDM_B(x, y)$ as follows, while ensuring balanced data distribution.

Definition 10 Let $OLDS^{\geq} = \left(U = U^l \cup U^u, A = C \cup D, V, f\right)$, $C = \{a_1, a_2, ..., a_l\}$ be an ordered high-dimensional partially labelled decision system, where l is the total number of features in C. For each $x, y \in U$ and $a_{l'} \in B \subseteq C$ $\left(l' \in \{1, 2, .., l\}\right)$ the distance between x and y w.r.t. the

feature a and the Euclidean distance metric (EDM) on feature $a_{l'}$ is defined as

$$d_a(x, y) = \begin{cases} \frac{f(x,a)-f(y,a)}{\max_a - \min_a}, & f(x,a) \geq f(y,a) \\ 0, & else \end{cases} \quad (19)$$

$$EDM_B(x, y) = \begin{cases} \sqrt{\frac{1}{l}\sum_{l'=1}^{l'} d_{a_{l'}}^2(x, y)}, & x \neq y \\ 0, & x = y \end{cases} \quad (20)$$

Correspondingly, Propositions 5 and 6 similarly need to be changed along with the new $EDM_B(x, y)$ matrix computation.

The information within high-dimensional ordered data can be harnessed more effectively through Definition 10. Although it introduces some level of error in extreme cases, these errors are deemed acceptable within a balanced dataset.

4.3 An accelerating semi-supervised feature selection algorithm for ordered partially labelled data based on neighbourhood discernibility degree with pseudolabel granular balls and distance matrix updating methods

By introducing the granular balls and the distance matrix updating methods, we can solve the following problems:

(1) When the number of samples in each category of the dataset is unbalanced, important features of different categories may be unfairly considered;
(2) The algorithm can be very time-consuming when dealing with large samples and high-dimensional datasets;

When using pseudolabel granular balls, a subset of samples with relatively balanced distributions can be obtained from the unbalanced samples by adjusting the threshold value of granular balls. And when facing high-dimensional datasets, since our algorithm is to judge one by one whether the features can be added to the candidate feature subset. Using distance matrix updating methods allows us to use the result of the previous calculation, reducing a lot of repeated calculations and thus saving a lot of time. Therefore, based on Proposition 4, a new semi-supervised feature selection algorithm for partially labelled ordered datasets is constructed by introducing the granular ball and distance matrix updating methods.

A problem is encountered when candidate features are selected by granular balls using integrated voting - granular balls containing different sample sizes vote with the same importance. For example, a GB_j containing 1000 samples and a GB_k containing 10 samples are both equally important in selecting candidate features. This is clearly not the result we want. Therefore, we redefined the weights of the features selected by each granular ball based on the number of samples contained in it. In this paper, we use the $SIG_U^{out}(b, B, D)$ defined in Definition 7 and the set of pseudolabel granular balls defined in Definition 8 to construct an importance weight belonging to each feature. Specific definitions are as follows.

Definition 11 Given an ordered partially labelled decision system $OLDS^{\geq} = \left(U = U^l \cup U^u, A = C \cup D, V, f\right)$, let $B = \emptyset$, $b \in C$, let $GB_B^U = \{GB_1, GB_2, \ldots, GB_n\}$ be the collection of all granular balls induced by B under $U, \forall GB_i \in GB_B^U (i \in [1, n])$, the importance weight belonging to each feature b can be defined as:

$$\omega(b, B, D, GB_i) = \frac{SIG_{GB_i}^{out}(b, B, D) \cdot |GB_i|}{|U|} \quad (21)$$

By Definition 11, we can use the mechanism of integrated voting more rationally to select candidate features. The detailed description of the feature selection algorithm based on the neighborhood discernibility degree using the pseudolabel granular ball selector and distance matrix updating methods is presented in Algorithm 2.

Algorithm 2 Neighbourhood Discernibility Degree Based Feature Selection Algorithm via Pseudolabel Granular Ball Selector and Distance Matrix Updating Methods (Algorithm NGM).

Input: An ordered partially labelled decision system $OLDS^{\geq} = (U = U^l \cup U^u, A = C \cup D, V, f)$ and neighbourhood radius δ;
Output: A feature subset Red.

Let $Red \leftarrow \emptyset$;
Calculate pseudolabel granular balls using the PGA algorithm;
Select the pseudolabel granular balls set $GB_C^U = \{GB_1, GB_2, \ldots, GB_n\}$;
Compute $NDD^{U^l}(C, D)$ and $NDD^{U^u}(C)$;
repeat
  A set of candidate features $AT \leftarrow \emptyset$;
  for each $i \in [1, n]$ **do**
    $\forall a \in C - Red$, compute $SIG_{GB_i}^{out}(a, Red, D)$;
    Select the feature b that satisfies $SIG_{GB_i}^{out}(b, Red, D) = \max\{SIG_{GB_i}^{out}(a, Red, D) : \forall a \in C - Red\}$;
    Let $\omega_b \leftarrow \frac{SIG_{GB_i}^{out}(b, Red, D) \cdot |GB_i|}{|U|}$, add $\{b : \omega_b\}$ into AT;
  end for
  Summing the ω_b of the same feature in AT;
  Select the feature b with the maximal ω_b in AT;
  $Red \leftarrow Red \cup \{b\}$;
  Compute $NDD^{U^l}(Red, D)$ and $NDD^{U^u}(Red)$ based on Proposition 5;
until $NDD^{U^l}(Red, D) \leq NDD^{U^l}(C, D)$ and $NDD^{U^u}(Red) \geq NDD^{U^u}(C)$
for $\forall a \in Red$ **do**
  Compute $SIG_{in}^U(a, Red, D)$ based on Proposition 6;
  if $SIG_{in}^U(a, Red, D) = 0$ **then**
    $Red \leftarrow Red - \{a\}$
  end if
end for
return Red

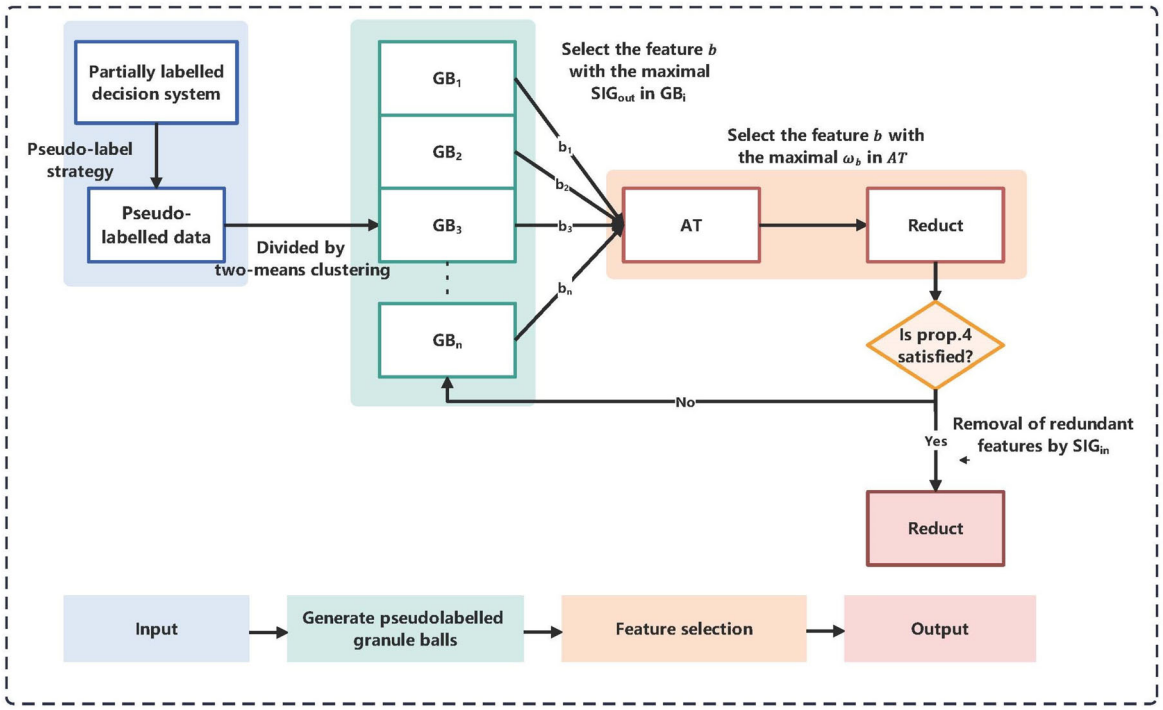


Fig. 2 NGM algorithm framework

The specific framework of the new algorithm is shown in Fig. 2. (1) label the partially labelled dataset via pseudo-label strategy and then divide the dataset by using two-means clustering to generate pseudolabel granular balls, (2) find the feature b that has the largest SIG_{out} value in each GB_i add it to AT, (3) assign weights based on the number of samples contained in each GB and select candidate features by integrated voting, (4) determine whether the subset of candidate features has satisfied the constraints, if so remove the redundant features through SIG_{in} and output the results, otherwise, return to (2).

Step 2-3 calculate pseudolabel granular balls by the Algorithm 1, whose time complexity is $O(K \times |U| \times |C|)$. To obtain $NDD^{U^l}(C, D)$ and $NDD^{U^u}(C)$, we need to calculate the distance between two objects which costs $O\left(|U|^2|C|\right)$. Steps 5-16 iterate through all the granular balls, selecting the most important features of each granular ball. This is achieved by integrating the voting form and weighting it according to the number of samples. The corresponding most important features are then selected and added to the Red until the termination condition is met. The time complexity of this process is $O\left(\sum_{i=1}^{n}|GB_i|^2\left(|C|+|C|-1+\ldots+1\right)\right) = O\left(\frac{\sum_{i=1}^{n}|GB_i|^2\left(|C|^2+|C|\right)}{2}\right)$.

Similarly, steps 19-24 is to delete the redundant features in Red and the time complexity is $O\left(\frac{\sum_{i=1}^{n}|U|^2\left(|Red|^2+|Red|\right)}{2}\right)$.

Consequently, the time complexity of the NGM algorithm is $O\left(\frac{\sum_{i=1}^{n}|GB_i|^2\left(|C|^2+|C|\right)}{2}\right)$. The asymptotic time complexity of NGM is $O\left(\sum_{i=1}^{n}|GB_i|^2|C|^2\right)$.

5 Incremental feature selection mechanism based on pseudolabel granular balls under multi-feature adding with ordered partially labelled data

In our daily life, the features of a sample change over time, and it is time-consuming to import all the features of the sample into the algorithm and recalculate them each time. Therefore, in this section we propose feature incremental algorithms to cope with these situations, which can use a subset of features obtained from previous feature selection combined with new feature samples for further selection.

When new sample features are added, a crucial question arises: how do we compare the importance of the newly added features with the original features that have already been screened for sample classification? Addressing this question assists us in identifying a relative globally optimal subset of features in the incremental algorithm. With Definition 11, we can easily compare the importance of each feature for classification, so when a new feature is added to a sample, we can first compare the importance of the new feature with that of the original feature. Subsequently, we remove the features in

Algorithm 3 Incremental Feature Selection Algorithm based on Neighbourhood Discernibility Degree via pseudolabel Granular Ball Selector and Distance Matrix Updating Methods when adding feature (Algorithm INGM-A).

Input:(1) An ordered partially labelled decision system $OLDS^{\geq} = (U = U^l \cup U^u, A = C \cup D, V, f)$ where $C = \{a_1, a_2, \ldots, a_l\}$, $C^+ = \{a_{l+1}, a_{l+2}, \ldots, a_{l+m}\}$and neighbourhood radius δ;
(2)The original reduced subset of features Red on U;
Output: A new reduct feature subset Red'.

Let $Red' \leftarrow Red$, $C' \leftarrow C \cup C^+$ and $Red \leftarrow \emptyset$;
Recalculate pseudolabel granular balls using the PGA algorithm;
Select the pseudolabel granular balls set $GB_C^U = \{GB_1, GB_2, \ldots, GB_n\}$;
A set of feature importance weights $W \leftarrow \emptyset$;
for each $i \in [1, n]$ **do**
  $\forall a \in C'$, compute $SIG^{out}_{GB_i}(a, Red, D)$
  Compute $\omega(a, Red, D, GB_i)$, add $\{a : \omega_a\}$ into W;
end for
Summing the ω_a of the same feature in W;
for $\forall a \in Red'$ **do**
  for $\forall a' \in C^+$ **do**
    if $\omega_a \leq \omega_{a'}$ **then**
      $Red' \leftarrow Red' - \{a\}$;
    end if
  end for
end for
Compute $NDD^{U^l}(Red', D)$ and $NDD^{U^u}(Red')$;
Compute $NDD^{U^l}(C', D)$ and $NDD^{U^u}(C')$;
while $NDD^{U^l}(Red', D) \leq NDD^{U^l}(C', D)$ and $NDD^{U^u}(Red') \geq NDD^{U^u}(C')$ **do**
  A set of candidate features $AT \leftarrow \emptyset$;
  for each $i \in [1, n]$ **do**
    $\forall a \in C' - Red'$, compute $SIG^{out}_{GB_i}(a, Red', D)$ based on Proposition 5;
    Select the feature b that satisfies $SIG^{out}_{GB_i}(b, Red', D) = \max\{SIG^{out}_{GB_i}(a, Red', D) : \forall a \in C' - Red'\}$;
    Compute $\omega(b, Red', D, GB_i)$, add $\{b : \omega_b\}$ into AT;
  end for
  Summing the ω_b of the same feature in AT;
  Select the feature b with the maximal ω_b in AT;
  $Red' \leftarrow Red' \cup \{b\}$;
  Compute $NDD^{U^l}(Red', D)$ and $NDD^{U^u}(Red')$;
end while
for $\forall a \in Red'$ **do**
  Compute $SIG^U_{in}(a, Red', D)$ based on Proposition 6;
  if $SIG^U_{in}(a, Red', D) = 0$ **then**
    $Red' \leftarrow Red' - \{a\}$;
  end if
end for
return Red'

the original feature subset that are deemed less important than the new feature, and then conduct feature selection again.

Based on Definition 11, an algorithm(INGM-A) for incremental feature selection when adding feature is proposed, which proceeds as follows.

In INGM-A algorithm, after the new features are added, the set of pseudolabel granule balls is first recalculated and generated. Then the importance of each feature is calculated and the features in the original feature subset with less importance than the added features are deleted. Afterward, the condition $NDD^{U^l}(Red', D) \leq NDD^{U^l}(C', D)$ and $NDD^{U^u}(Red') \geq NDD^{U^u}(C')$ is checked to determine if it is satisfied. If satisfied, the algorithm proceeds directly to deleting the redundant feature portion of the feature subset. If not satisfied, the new $(EDM_C)^U_M$ and $(NR_C)^U_M$ are computed using Proposition 5, allowing for the selection of the new feature with the largest $SIG^{out}_{GB_i}(a, Red', D)$ value. Subsequently, the process of removing redundant features is repeated.

Step 2-3 recalculate pseudolabel granular balls by the Algorithm 1, whose time complexity is $O(K \times |U| \times |C|)$. The time complexity of steps 4-16 can be approximated to be equal to the time complexity of computing the $SIG^{out}_{GB_i}(a, Red, D)$, thus it is $O\left(\sum_{i=1}^{n} |GB_i|^2 |C'|\right)$. In steps 17ÍC18, we need to calculate the $NDD^{U^l}(C', D)$ and $NDD^{U^u}(C')$, whose time complexity is $O\left(|U|^2|C'|\right)$. Then, steps 19ÍC30 is the process of filtering out representative features from the remaining features, whose worst case time complexity is $O\left(\frac{\sum_{i=1}^{n}|GB_i|^2\left(|C'-Red'|^2+|C'-Red'|\right)}{2}\right)$. Similarly, steps 31-36 is to delete the redundant features in Red, the time complexity is $O\left(\frac{\sum_{i=1}^{n}|U|^2\left(|Red'|^2+|Red'|\right)}{2}\right)$. Therefore, the time complexity of the INGM-A algorithm is $O\left(\frac{\sum_{i=1}^{n}|GB_i|^2\left(|C'-Red'|^2+|C'-Red'|\right)}{2}\right)$.

6 Experimental analysis

In this section, we will experiment with the previously proposed algorithm to test and evaluate its performance. The experimental results are reported and analyzed in detail. Extensive experiments are carried out on a PC with Windows 10, 16 GB of RAM and a 2.60 GHz i7-10750H CPU. The programming language is Python.

6.1 Datasets and experimental settings

There are 12 datasets used for experiments in this paper, seven of which are from the UCI machine learning repository [38], four datasets on scikit-feature[1] and one on Kaggle. Table 2 shows the details of these datasets.

Before using the dataset, each numerical feature of the dataset needs to be normalised to between 0 and 1. In the experiment, the dataset was divided into training and test sets in the ratio of 80% and 20%. In order to better validate the performance of the algorithm, this paper adopts a ten-

[1] Data sources: https://jundongl.github.io/scikit-feature/

Table 2 Description of the 12 datasets

ID	Datasets	Objects	Features	Classes
1	Wine	178	13	2
2	Wholesale customers	440	7	2
3	Pima Indians Diabetes	768	8	2
4	Breast Cancer(wdbc)	569	31	2
5	Image Segmentation	2310	18	7
6	Rice	3180	7	2
7	Gas Sensor Array-7	3613	128	6
8	Dry Bean	13611	16	7
9	warpAR10P	130	2400	10
10	warpPIE10P	210	2420	10
11	pixraw10P	100	10000	10
12	orlraws10P	100	10304	10

fold cross-validation approach and chooses three classifiers, SVM, KNN and Random Forest to examine the classification accuracy of the selected features. The kernel function of the SVM is a linear kernel, the k value of the KNN is set to 3, and the number of trees in the Random Forest model is set to 100. The following experimental predictive performance is validated by the average accuracy across the 10 folds.

6.2 Baselines

To validate the efficiency of the proposed NGM and INGM-A algorithm, four classic feature selection algorithms are compared, including supervised, semi-supervised and incremental attribute ones. All methods concerned are briefly introduced as follows.

(1) SFSS [39]: A supervised fast and robust feature selection based on the separability in fuzzy decision systems.
(2) HAR-A [20]: An incremental feature selection algorithm based on matrix and using conditional entropy for ordered data set with time-evolving features.
(3) SemiFREE [24]: A semi-supervised feature selection with fuzzy relevance and redundancy.
(4) FScNCE [25]: A supervised heterogeneous feature selection based on neighborhood combination entropy.

The following experiments are divided into four aspects:

(i) Comparison of algorithm performance at different pseudolabel granular ball purity thresholds.
(ii) Comparison of classification accuracy of NGM algorithm with different neighbourhood radius δ and labelled data rate.
(iii) Comparisons among the NGM and INGM-A algorithm and the other four algorithms at different labelling ratios.
(iv) Statistical comparison of NGM, INGM-A algorithms and four other algorithms on multiple datasets.

6.3 Comparison of algorithm performance at different pseudolabel granular ball purity thresholds

From the Algorithm 1, it is known that the number of granule ball generated is closely related to the purity threshold. And the number of granule ball is an important factor that affects the calculation time of the algorithm. Thus, the primary goal in this subsection is to find a relatively optimal purity threshold that allows the algorithm to have a high performance without loss of accuracy.

When the purity threshold is too small, the number of granule ball and the purity of the samples inside them will be low; when the purity threshold is too high, there will be very few samples inside certain granule ball. Therefore, we will set $\{0.5, 0.6, 0.7, 0.8\}$ as the purity thresholds for each dataset to see its impact on the performance of the Algorithm 2. 10% of the data in each dataset were randomly selected to retain their labels as labelled data, with the rest designated as unlabelled data and the current relatively optimal neighbourhood radius δ is selected for each dataset. We tested the selected subset of features on SVM,KNN and Random Forest respectively. Since the overall trends of the three classifiers are roughly the same, only the trend of classification accuracy of SVM with purity threshold is shown. The results are shown in Table 3. The following conclusions can be drawn from the experimental data:

(1) The pseudolabel granular ball strategy did not work well on the binary classification dataset. Mostly, the granule ball purity reached 1 after the first split, so the number of granule ball was 2.
(2) In most cases, the higher the purity threshold and the greater the number of granule ball, the shorter the computational time of the algorithm. This is also consistent with the time complexity of the Algorithm 2 obtained in Section 4, indicating that the pseudolabel granular ball strategy can indeed reduce the computational time of the feature selection algorithm.
(3) Most of the optimal purity thresholds are found to be distributed between 0.6 and 0.7. Setting the purity threshold too high can actually result in decreased classification accuracy. This may be due to poor feature selection because some high-purity granule balls contain too small a sample size.

Table 3 SVM Classification accuracy (%) and computational time in relation to the four purity threshold (Better values are in bold)

Dataset	Radius δ	Purity Threshold	GB num	SVM Accuracy (%)	Time (s)
Wine	0.10	0.50	2	98.30	2.99
		0.60	2	98.30	2.99
		0.70	3	**98.30**	**2.56**
		0.80	3	98.30	2.99
Wholesale customers	0.007	0.50	2	87.27	5.56
		0.60	2	87.27	5.56
		0.70	2	**87.27**	**5.56**
		0.80	2	87.27	5.56
Diabetes	0.005	0.50	2	77.09	22.46
		0.60	2	77.09	22.46
		0.70	2	**77.09**	**22.46**
		0.80	2	77.09	22.46
Wdbc	0.02	0.50	2	96.48	66.08
		0.60	2	96.48	66.08
		0.70	2	**96.48**	**66.08**
		0.80	2	96.48	66.08
Segmentation	0.012	0.50	5	90.65	311.00
		0.60	6	90.65	284.30
		0.70	7	**90.65**	**256.90**
		0.80	37	90.65	280.00
Rice	0.003	0.50	2	92.89	570.00
		0.60	2	92.89	570.00
		0.70	2	**92.89**	**570.00**
		0.80	2	92.89	570.00
Batch	0.01	0.50	5	96.40	7417.00
		0.60	5	**96.40**	**7417.00**
		0.70	6	95.27	5990.00
		0.80	31	95.55	4033.00
Dry Bean	0.001	0.50	5	89.47	7637.00
		0.60	7	89.47	6462.00
		0.70	7	**89.47**	**6462.00**
		0.80	7	89.47	6462.00
warpAR10P	0.01	0.50	6	92.31	691.14
		0.60	6	**92.31**	**691.14**
		0.70	10	88.46	469.02
		0.80	10	88.46	469.02
warpPIE10P	0.0025	0.50	6	95.24	1158.28
		0.60	10	97.14	976.55
		0.70	16	**97.14**	**768.25**
		0.80	21	97.14	769.57
pixraw10P	0.008	0.50	6	85.00	474.15
		0.60	9	**97.00**	**499.23**
		0.70	10	89.00	417.20
		0.80	10	89.00	417.20

Table 3 continued

Dataset	Radius δ	Purity Threshold	GB num	SVM Accuracy (%)	Time (s)
orlraws10P	0.004	0.50	7	88.00	622.21
		0.60	9	88.00	546.24
		0.70	12	**94.00**	**552.23**
		0.80	13	88.00	520.22

The best results are highlighted in bold and underline entities in these tables

6.4 Comparison of classification accuracy with different neighbourhood radius and labelled data rate

In this subsection, we explore the effect of NGM on classification accuracy for different neighbourhood radius δ and different labelled data rates. Our objective is to determine the optimal neighborhood radius δ and optimal labelled data rate for each dataset. {10%, 20%, 30%, 40%} of the data in each dataset were randomly selected to retain their labels as labelled data, with the rest designated as unlabelled data.

We roughly determine an optimal neighborhood radius and then select some other radius values around that radius to observe how it affects the classification accuracy. Since the overall trends of the three classifiers SVM, KNN, and Random Forest are not substantially different from each other, the following only presents trend graphs illustrating the classification accuracy of SVM with varying neighborhood radius and labelled data rate.

Analysing Fig. 3 we can get the following conclusion:

(1) For most datasets, the optimal neighbourhood radius of the NGM algorithm is between [0.001, 0.009].
(2) In most datasets, the NGM algorithm tends to perform better with low labelled data rates. In Fig. 3(d) and (e), NGM provides the highest classification accuracy using only 10% labelled data rate. In Fig. 3(j) and (k), NGM provides the highest classification accuracy using only 20% labelled data rate. This confirms the viability of the pseudolabel granular balls method for selecting features on datasets with low label rates.

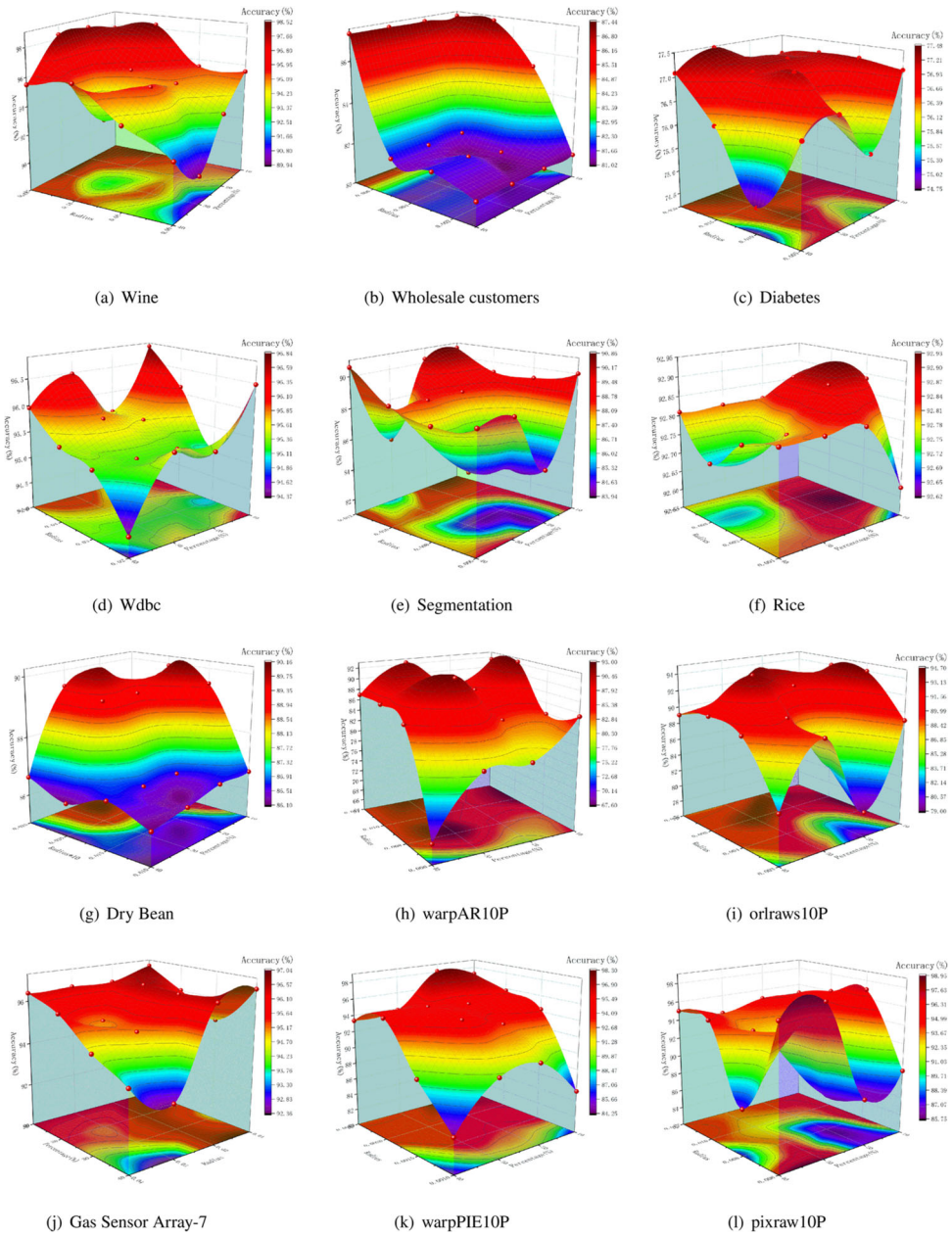
6.5 Comparison with other algorithms

In this section, we evaluate the performance of the proposed algorithm NGM and INGM-A compared with the other four algorithms in terms of time consumption and classification accuracy under different labelling ratios. For the INGM-A algorithm, we substitute the last 20% of all features of each dataset as added features to be calculated in the algorithm. Substituting the relative optimal neighbourhood radius δ (Given in Table 4) obtained in the previous subsection for different labelled data rates into the NGM and INGM-A algorithms for comparison with the other algorithms. Since supervised algorithms cannot handle unlabelled data, for supervised algorithms they use only labelled data in the dataset for feature selection, i.e., {10%, 20%, 30%, 40%} samples in the dataset, respectively. The results can be obtained as shown in the Tables 4, 5, 6 and 7.

Tables 4-6 demonstrates the ten-fold cross validation classification accuracy means for different algorithms with labelled data rates of {10%, 20%, 30%, 40%} under the three classifiers SVM, KNN and Random Forest. In addition, the size of the feature subset selected by each algorithm is provided in Table 7. From the experimental data we can draw the following conclusions.

(1) The NGM and INGM-A algorithms have better classification accuracies compared to the other algorithms in most cases when the number of feature subsets being selected is similar.
(2) In most cases, the classification accuracies and the size of the feature subsets of the INGM-A and NGM algorithms are not very different, so the experimental results provide strong evidence that Definition 11 is useful for incremental algorithms to search for globally optimal solutions.
(3) The NGM and INGM-A algorithms have a clear advantage over other classifiers for classification on KNN. This is because the pseudolabel granular balls are generated by the K-means method, causing the algorithms to screen out features that are more suitable for KNN classifiers. Therefore, if you want to improve the classification accuracy of the dataset on KNN, the NGM and INGM-A algorithms can be good choices.
(4) The NGM and INGM-A algorithms have extremely obvious advantages in high-dimensional low labeling rate datasets such as $pixraw10P, orlraws10P$. This goes some way to prove that distance functions and matrix updating methods changed for high-dimensional data in Definition 10 is fruitful.

Table 8 visually compares the six algorithms at the time-consuming level. Since the algorithm proposed in this article is a semi-supervised feature selection algorithm, it is not of the same order of magnitude as the datasets run by most of the other supervised feature selection algorithms. Therefore, in Table 8, only the runtimes for six large sample or high-dimensional datasets with a 40% labelled data rate are shown.



(a) Wine (b) Wholesale customers (c) Diabetes

(d) Wdbc (e) Segmentation (f) Rice

(g) Dry Bean (h) warpAR10P (i) orlraws10P

(j) Gas Sensor Array-7 (k) warpPIE10P (l) pixraw10P

Fig. 3 Classification accuracy of NGM algorithm under SVM with different neighbourhood radius and labelled data rate

If the semi-supervised feature selection algorithm runs on a dataset of size $|U|$, then the supervised feature selection algorithm runs on a dataset of size $40\%|U|$. As can be seen from Table 7, the HAR-A algorithm has little or no selection ability for some high-dimensional datasets, and in most cases only one feature can be selected. Therefore we do not show its corresponding dataset running time in Table 8. By comparing the data in the table we can get the following conclusions.

(1) At the time complexity level, the NGM and INGM-A algorithms have a clear advantage over the other three algorithms except SFSS.

(2) Our proposed algorithm saves nearly six times the time compared to SemiFREE, which is also a semi-supervised feature selection algorithm, when running the $DryBean$ dataset with a large amount of sample data. Compared to the supervised feature selection algorithm(except SFSS), the running time of our algorithm is also not much different. Therefore, this experimental data is a good validation that pseudolabel granular balls are effective in saving the running time of large sample size datasets.

(3) On high-dimensional datasets such as $warpPIE10P$ and $pixraw10P$ with a large number of features, our algorithm still has a significant lead over SemiFree. However, comparing to other supervised feature selection

Table 4 SVM Classification accuracy (%) in relation to the four labelling rates (Better values are in bold)

Dataset	Radius δ	Percentage	Algorithm					
			NGM	INGM-A	SFSS	SemiFREE	HAR-A	FScNCE
Wine	0.10	10.00%	**98.30**	98.30	95.56	97.19	96.67	96.11
	0.12	20.00%	**98.30**	98.30	97.22	97.19	97.78	98.30
	0.12	30.00%	**98.30**	98.30	97.22	97.19	97.19	96.08
	0.10	40.00%	96.05	96.60	94.44	96.05	**97.22**	96.60
Wholesale Customers	0.007	10.00%	**87.27**	87.27	86.82	87.27	87.27	86.82
	0.007	20.00%	**87.27**	87.27	86.82	87.27	85.23	86.82
	0.007	30.00%	**87.27**	87.27	86.82	87.27	87.27	87.27
	0.007	40.00%	**87.27**	87.27	86.82	87.27	87.27	87.27
Diabetic	0.005	10.00%	77.09	**77.21**	76.68	76.82	76.69	76.69
	0.015	20.00%	77.08	**77.21**	75.63	76.82	76.69	76.69
	0.020	30.00%	**77.47**	77.21	76.41	76.82	76.69	75.26
	0.020	40.00%	**77.21**	77.21	76.41	76.82	76.69	76.69
Wdbc	0.020	10.00%	**96.48**	96.13	93.33	93.67	95.95	95.60
	0.040	20.00%	95.60	95.60	93.68	93.67	96.31	**96.66**
	0.050	30.00%	96.48	96.48	93.68	93.67	**96.84**	96.13
	0.050	40.00%	95.96	96.13	93.68	93.67	**96.49**	95.96
Segmentation	0.012	10.00%	**90.65**	89.31	80.00	86.36	83.38	86.97
	0.012	20.00%	**90.26**	90.04	80.00	81.26	87.32	88.27
	0.010	30.00%	88.44	90.04	80.00	81.26	**90.43**	88.23
	0.012	40.00%	**90.56**	90.04	80.00	81.26	87.32	90.13
Rice	0.002	10.00%	**92.89**	92.89	92.68	92.68	92.73	92.73
	0.002	20.00%	**92.89**	92.89	92.68	92.70	92.60	92.70
	0.002	30.00%	92.78	92.70	92.83	92.70	92.60	**92.89**
	0.002	40.00%	92.78	92.70	92.68	92.70	**92.97**	92.89
Gas Sensor Array-7	0.030	10.00%	96.40	96.40	92.03	65.95	88.74	**98.51**
	0.020	20.00%	96.54	95.16	93.25	73.45	43.29	**99.50**
	0.020	30.00%	96.46	95.10	93.25	73.45	–	**99.64**
	0.010	40.00%	96.76	92.50	93.25	81.53	–	**99.78**
Dry Bean	0.015	10.00%	84.26	84.04	84.20	84.11	**85.25**	84.26
	0.002	20.00%	83.46	83.78	84.20	84.11	**85.26**	83.78
	0.0025	30.00%	84.26	84.04	84.20	84.11	**85.36**	84.04
	0.0025	40.00%	84.27	84.04	84.20	84.11	**85.48**	84.04
warpAR10P	0.010	10.00%	**89.64**	88.89	88.01	83.70	83.70	89.39
	0.010	20.00%	90.12	88.89	87.53	83.70	85.80	**91.36**
	0.008	30.00%	90.74	88.89	88.01	83.70	85.80	**92.10**
	0.008	40.00%	**91.23**	88.89	88.01	83.70	85.80	89.63
warpPIE10P	0.0025	10.00%	87.07	87.07	83.89	86.04	86.04	**88.78**
	0.0025	20.00%	87.56	87.07	83.89	86.04	86.04	**88.29**
	0.0020	30.00%	**88.29**	87.07	83.89	86.04	86.04	87.32
	0.0020	40.00%	**89.27**	87.07	83.89	86.04	86.04	88.05
pixraw10P	0.008	10.00%	76.00	77.00	74.00	74.50	73.50	**78.00**
	0.010	20.00%	76.50	77.50	75.50	74.50	75.50	**78.50**

Table 4 continued

Dataset	Radius δ	Percentage	Algorithm					
			NGM	INGM-A	SFSS	SemiFREE	HAR-A	FScNCE
	0.006	30.00%	77.00	77.00	76.50	74.50	75.50	**79.00**
	0.006	40.00%	77.50	78.00	77.50	74.50	75.50	**79.50**
orlraws10P	0.004	10.00%	**78.50**	77.00	77.50	74.50	74.50	75.50
	0.005	20.00%	79.00	**79.00**	77.50	74.50	74.50	75.50
	0.005	30.00%	**79.50**	79.00	78.50	74.50	74.50	75.50
	0.005	40.00%	**80.00**	79.00	79.00	74.50	74.50	75.50

The best results are highlighted in bold and underline entities in these tables

Table 5 KNN Classification accuracy (%) in relation to the four labelling rates (Better values are in bold)

Dataset	Percentage	Algorithm					
		NGM	INGM-A	SFSS	SemiFREE	HAR-A	FScNCE
Wine	10.00%	**97.75**	97.75	93.82	94.97	93.24	93.86
	20.00%	**97.75**	97.75	95.56	94.97	96.08	96.60
	30.00%	**97.75**	97.75	95.56	94.97	95.49	94.97
	40.00%	96.60	**96.63**	96.11	92.68	96.08	94.93
Wholesale Customers	10.00%	**91.14**	87.50	89.55	91.14	87.50	87.27
	20.00%	**91.14**	87.50	89.55	91.14	87.50	87.27
	30.00%	**91.14**	87.50	89.55	91.14	87.50	87.50
	40.00%	**91.14**	87.50	89.55	91.14	87.50	87.50
Diabetic	10.00%	**76.18**	74.74	75.63	73.70	73.57	73.44
	20.00%	74.74	74.47	**75.77**	73.70	73.57	73.57
	30.00%	73.96	**74.47**	72.90	73.70	73.57	73.30
	40.00%	**74.74**	74.74	72.90	73.70	73.57	73.57
Wdbc	10.00%	95.07	**95.60**	93.33	94.55	95.25	94.71
	20.00%	**96.13**	96.13	93.86	94.56	94.72	95.77
	30.00%	95.60	95.60	93.86	94.55	**96.12**	95.95
	40.00%	**95.95**	94.72	93.86	94.55	95.77	94.90
Segmentation	10.00%	94.03	93.51	88.74	**94.63**	90.87	91.47
	20.00%	**93.59**	93.51	88.74	88.83	91.69	92.99
	30.00%	93.16	**93.38**	88.74	88.83	92.29	92.55
	40.00%	**94.03**	93.38	88.74	88.83	91.69	92.03
Rice	10.00%	92.47	92.47	92.28	92.55	92.13	**92.78**
	20.00%	**92.47**	92.47	92.28	92.18	92.13	92.36
	30.00%	**92.57**	92.36	91.99	92.18	92.13	92.47
	40.00%	**92.57**	92.36	92.28	92.18	92.15	92.47
Gas Sensor Array-7	10.00%	**99.00**	99.00	99.14	88.15	97.65	98.42
	20.00%	99.09	98.31	99.20	93.55	41.96	**99.47**
	30.00%	98.81	99.11	99.20	93.55	–	**99.47**
	40.00%	99.25	98.39	99.20	96.02	–	**99.47**
Dry Bean	10.00%	**88.44**	82.60	85.20	85.87	74.49	85.19
	20.00%	**88.17**	82.60	85.20	85.87	74.49	87.64
	30.00%	88.73	82.60	85.20	85.87	**89.22**	87.64
	40.00%	85.13	82.60	85.20	85.87	**89.12**	87.64

Table 5 continued

Dataset	Percentage	Algorithm					
		NGM	INGM-A	SFSS	SemiFREE	HAR-A	FScNCE
warpAR10P	10.00%	54.62	**59.23**	53.85	43.85	20.00	49.23
	20.00%	**53.08**	50.77	50.77	41.54	20.00	41.54
	30.00%	**65.38**	53.85	50.00	40.77	20.00	50.77
	40.00%	50.00	50.00	52.31	42.31	20.00	**69.23**
warpPIE10P	10.00%	81.43	83.33	87.14	**89.52**	19.05	84.76
	20.00%	84.29	85.71	88.57	**88.57**	19.05	88.57
	30.00%	84.76	84.76	80.00	**85.24**	19.05	85.24
	40.00%	74.29	74.29	81.43	86.67	19.05	**91.43**
pixraw10P	10.00%	**86.00**	86.00	–	72.00	68.00	70.00
	20.00%	88.00	88.00	80.00	70.00	68.00	**96.00**
	30.00%	84.00	**89.00**	82.00	70.00	68.00	84.00
	40.00%	82.00	81.00	88.00	78.00	68.00	**95.00**
orlraws10P	10.00%	80.00	**85.00**	24.00	77.00	59.00	65.00
	20.00%	80.00	**87.00**	79.00	65.00	59.00	80.00
	30.00%	**84.00**	82.00	80.00	57.00	59.00	82.00
	40.00%	86.00	79.00	83.00	57.00	59.00	**87.00**

The best results are highlighted in bold and underline entities in these tables

Table 6 Random Forest Classification accuracy (%) in relation to the four labelling rates (Better values are in bold)

Dataset	Percentage	Algorithm					
		NGM	INGM-A	SFSS	SemiFREE	HAR-A	FScNCE
Wine	10.00%	96.11	96.11	**97.78**	96.67	96.67	96.08
	20.00%	96.11	96.11	**98.33**	96.67	98.33	97.22
	30.00%	96.11	96.11	**98.33**	96.67	97.22	96.08
	40.00%	**98.33**	96.11	97.78	96.67	98.33	96.63
Wholesale Customers	10.00%	91.82	90.91	91.59	91.82	90.91	**92.27**
	20.00%	91.82	90.91	91.59	91.82	92.05	**92.27**
	30.00%	**91.82**	90.91	91.59	91.82	90.91	90.91
	40.00%	**91.82**	90.91	91.59	91.82	90.91	90.91
Diabetic	10.00%	74.99	74.48	75.90	75.78	**75.91**	74.87
	20.00%	74.48	74.48	**76.29**	75.78	75.91	75.91
	30.00%	**77.73**	74.48	76.42	75.78	75.91	75.00
	40.00%	74.48	74.48	**76.42**	75.78	75.91	75.91
Wdbc	10.00%	**96.13**	95.43	94.38	94.38	96.13	95.78
	20.00%	95.26	95.26	94.38	94.73	**97.01**	95.78
	30.00%	96.13	96.13	94.38	94.38	**96.84**	95.44
	40.00%	95.26	96.31	94.38	94.38	**96.66**	95.08
Segmentation	10.00%	97.36	**97.45**	91.52	97.14	96.32	96.32
	20.00%	97.27	97.66	91.52	92.77	97.45	**97.79**
	30.00%	97.19	97.32	91.52	92.77	**97.40**	97.32
	40.00%	97.32	97.32	91.52	92.77	**97.45**	97.32
Rice	10.00%	**92.20**	92.20	92.20	91.68	91.78	92.02
	20.00%	**92.20**	92.10	92.20	91.50	91.73	91.57

Table 6 continued

Dataset	Percentage	Algorithm					
		NGM	INGM-A	SFSS	SemiFREE	HAR-A	FScNCE
	30.00%	**92.26**	91.57	92.07	91.50	91.73	92.20
	40.00%	**92.26**	91.57	92.20	91.50	91.97	92.20
Gas Sensor Array-7	10.00%	**99.34**	99.34	99.28	87.07	97.73	99.00
	20.00%	99.39	98.87	99.45	93.99	37.72	**99.81**
	30.00%	**99.47**	98.87	99.45	93.99	–	99.28
	40.00%	**99.45**	97.65	99.45	96.18	–	99.36
Dry Bean	10.00%	**86.53**	59.45	76.34	68.72	67.14	80.69
	20.00%	84.63	59.45	76.34	68.72	67.14	**85.23**
	30.00%	**86.04**	59.45	76.34	68.72	83.88	85.23
	40.00%	80.26	59.45	76.34	68.72	83.58	**85.23**
warpAR10P	10.00%	83.85	**86.15**	78.46	51.54	16.15	73.85
	20.00%	81.54	**86.92**	80.00	52.31	16.15	82.31
	30.00%	**86.15**	82.31	74.62	50.00	16.15	82.31
	40.00%	79.23	79.23	76.15	50.77	16.15	**83.85**
warpPIE10P	10.00%	**95.71**	95.71	93.33	94.29	22.38	94.29
	20.00%	**97.62**	97.14	95.24	95.24	22.38	95.71
	30.00%	96.19	96.19	92.86	94.76	22.38	**96.67**
	40.00%	96.19	96.19	93.81	93.81	22.38	**99.52**
pixraw10P	10.00%	**100.00**	97.00	–	88.00	71.00	98.00
	20.00%	**100.00**	96.00	98.00	89.00	71.00	98.00
	30.00%	98.00	**99.00**	93.00	82.00	71.00	98.00
	40.00%	**99.00**	98.00	98.00	91.00	71.00	98.00
orlraws10P	10.00%	**97.00**	86.00	16.00	87.00	50.00	92.00
	20.00%	**96.00**	90.00	91.00	91.00	50.00	95.00
	30.00%	**94.00**	93.00	93.00	80.00	50.00	93.00
	40.00%	91.00	89.00	90.00	81.00	50.00	**95.00**

The best results are highlighted in bold and underline entities in these tables

Table 7 Feature subset size in relation to the four labelling rates

Dataset	Raw data	Percentage	Algorithm					
			NGM	INGM-A	SFSS	SemiFREE	HAR-A	FScNCE
Wine	13	10.00%	9	9	11	10	10	9
		20.00%	9	9	10	10	10	12
		30.00%	9	9	10	10	11	11
		40.00%	8	10	11	11	11	11
Wholesale customers	7	10.00%	5	7	7	6	6	6
		20.00%	6	7	7	6	5	6
		30.00%	6	7	7	6	7	7
		40.00%	5	7	7	6	7	7
Diabetic	8	10.00%	5	5	5	7	8	7
		20.00%	5	5	5	7	8	8
		30.00%	5	6	5	7	8	6
		40.00%	5	6	5	7	8	8

Table 7 continued

Dataset	Raw data	Percentage	Algorithm					
			NGM	INGM-A	SFSS	SemiFREE	HAR-A	FScNCE
Wdbc	31	10.00%	7	12	15	15	14	9
		20.00%	11	11	12	15	17	14
		30.00%	13	13	12	15	20	13
		40.00%	12	13	12	15	23	12
Segmentation	18	10.00%	8	7	10	9	12	7
		20.00%	7	8	10	12	13	8
		30.00%	9	8	10	12	14	9
		40.00%	7	8	10	12	13	10
Rice	7	10.00%	4	4	4	4	3	4
		20.00%	4	4	4	4	4	4
		30.00%	3	4	5	4	4	4
		40.00%	4	4	4	4	6	4
Gas Sensor Array-7	128	10.00%	12	12	10	12	5	13
		20.00%	12	11	11	12	1	21
		30.00%	12	11	11	12	1	27
		40.00%	12	12	11	12	1	33
Dry Bean	16	10.00%	5	4	7	7	3	4
		20.00%	5	4	7	7	3	14
		30.00%	5	4	7	7	10	13
		40.00%	5	4	7	7	11	14
warpAR10P	2400	10.00%	51	60	55	50	1	15
		20.00%	48	55	40	43	1	48
		30.00%	35	49	40	43	1	52
		40.00%	34	34	35	43	1	33
warpPIE10P	2420	10.00%	36	32	40	36	1	35
		20.00%	38	38	30	38	1	27
		30.00%	29	29	30	32	1	23
		40.00%	27	27	30	32	1	67
pixraw10P	10000	10.00%	14	15	–	16	1	11
		20.00%	13	16	16	16	1	26
		30.00%	13	15	16	16	1	13
		40.00%	13	16	16	16	1	16
orlraws10P	10304	10.00%	25	23	26	25	1	20
		20.00%	25	29	25	25	1	18
		30.00%	26	26	25	25	1	30
		40.00%	22	26	23	25	1	42

Table 8 Time-consuming comparison of six algorithms(s)(Better values are in bold)

Algorithm	Time Complexity	Dataset(40%)					
		Gas-7	Dry Bean	warpPIE	warpAR	pixraw	orlraws
NGM	$O\left(\frac{\sum_{i=1}^{n}\lvert GB_i\rvert^2\left(\lvert C\rvert^2+\lvert C\rvert\right)}{2}\right)$	4974	6050	**549**	**704**	549	434
INGM-A	$O\left(\frac{\sum_{i=1}^{n}\lvert GB_i\rvert^2\left(\lvert C'-Red'\rvert^2+\lvert C'-Red'\rvert\right)}{2}\right)$	5398	3247	594	1054	634	730
SFSS	$O(\lvert C\rvert^2\lvert U\rvert\lvert U/D\rvert)$	**42**	**17**	856	1520	**545**	**374**
SemiFREE	$O(\lvert U\rvert^2\lvert C\rvert\lvert Red\rvert)$	29615	44727	4022	4782	5894	2887
HAR-A	$O\left(\lvert U\rvert\lvert C^+\rvert\lvert C'\rvert+(\lvert C'\rvert-\lvert Red'\rvert)\lvert U\rvert^2+\lvert B\rvert^2\lvert U\rvert^2\right)$	–	5403	–	–	–	–
FScNCE	$O(\lvert U\rvert^2\lvert C\rvert^2)$	11425	5743	4326	28296	7772	20530

The best results are highlighted in bold and underline entities in these tables

algorithms, our algorithm achieves a great reversal in terms of runtime. Even when comparing to the SFSS algorithm that only uses 40% of the sample size, our algorithm's runtime is comparable. This is strong evidence that the distance matrix updating method proposed in Section 4.2 is extremely effective in reducing the runtime for datasets with a large number of features.

6.6 Statistical comparisons over multiple datasets

In order to further analyse whether our proposed algorithm is feasible or not, in this subsection we validate the obtained results statistically. For this purpose, we introduce the Friedman test and BonferroniÍCDunn test [40], which are well known in the field of machine learning. The specific Friedman test expressions are as follows.

$$\chi_F^2 = \frac{12N}{K(K+1)\left[\sum_{r=1}^{K} R_r^2 - \frac{K(K+1)^2}{4}\right]} \tag{22}$$

$$F_F = \frac{(N-1)\chi_F^2}{N(K-1)-\chi_F^2} \tag{23}$$

where N represents the number of datasets, K represents the number of algorithms, and R_r represents the average ranking of the rth algorithm.

In the experiments of this article, each dataset was experimented with four labelling rates, and the classification accuracy of each dataset on each algorithm was obtained by averaging the classification accuracies of the four results of each dataset. This results in $K = 6$ and $N = 12$, and the Friedman test can be performed to obtain the results in Table 9.

As observed from Table 9, the F_F values for all three classifiers are greater than 2.38. Hence, at a significance level of $\alpha = 0.05$, the null hypothesis H_0, stating that there is no significant difference in the performance of all algorithms, can be well rejected. That means there is a significant difference between the algorithms.

In order to present a more intuitive picture of the differences between each algorithm, we performed a BonferroniÍCDunn test as a post hoc test. Whether there is a significant difference between the performance of the two algorithms can be measured by the following critical difference:

$$CD = q_\alpha\sqrt{\frac{K(K+1)}{6N}} \tag{24}$$

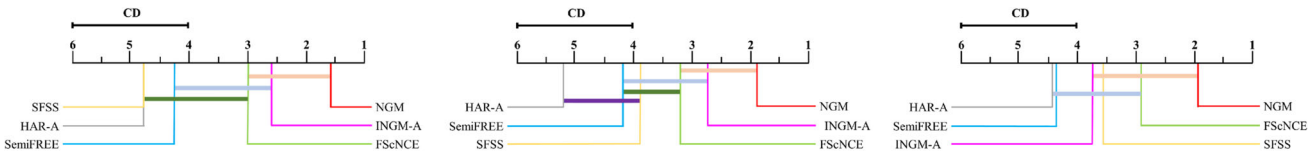
The value of q_α is 2.576 when $K = 6$, $N = 12$ and $\alpha = 0.05$, so by calculating we can get $CD = 1.967$. When the difference between the average rankings of two algorithms is greater than or equal to CD, it can be assumed that there is a significant performance difference between them.Visualising the results gives a CD diagrams as in Fig. 4.

From Fig. 4, the following conclusions can be drawn:

(1) The NGM algorithm has the highest average ranking among all the classifiers, which is a good indication that the NGM algorithm has a very good performance among these six algorithms, and the information contained in the features it selects is much richer.
(2) On the KNN classifier, the largest differences between algorithms were observed, with all algorithms being divided into four categories. The NGM and INGM-A algorithms remained in the top two positions. This fur-

Table 9 Classifier Friedman test and Critical values($\alpha = 0.05$)

Classifier	F_F	Critical Values($F(5, 55)$)
SVM	11.15	2.38
KNN	7.18	
Random Forest	3.66	



(a) SVM (b) KNN (c) Radom Forest

Fig. 4 Critical difference diagrams of average ranking of algorithms on different classifiers($\alpha = 0.05$)

ther shows that the pseudolabel granular balls approach is helpful in improving the classification accuracy of KNN classifiers.

(3) Statistically, the NGM algorithm is always significantly different from SemiFREE, which is also a semi-supervised feature selection algorithm. In terms of SVM and KNN, it is significantly different from HAR-A and SFSS too. At the same time, its performance is equivalent to that of FScNCE and INGM-A algorithms.

7 Conclusion and future work

In this paper, we propose a semi-supervised feature selection algorithm NGM applied on partially labelled ordered datasets. Meanwhile, an incremental algorithm INGM-A is proposed on the basis of the NGM algorithm for the increase of features over time. The novelty of the algorithms is that firstly they both use pseudolabel granular balls method to select more representative features by voting on each granular balls. Additionally, by the dividing of granular balls, they effectively reduce the computation time for large sample datasets. Secondly, they both use matrix update method to reduce the computation time for datasets with large number of features. Thirdly the INGM-A algorithm finds an approximate global optimal solution by comparing the importance of the newly added features with the original features. Through experiments and statistical analyses, it is evident that the NGM and INGM-A algorithms significantly reduce computation time compared to other algorithms while still selecting more informative features, resulting in a higher classification accuracy.

So far, we have only investigated algorithms for the selection of numerical class features. In the future, we will further refine our current work to deal with datasets with categorical features and features with missing values. In addition, we will study other feature measures such as fuzzy relevance, conditional entropy, etc., to select the optimal feature subset by the granular ball and matrix integration voting methods proposed in this paper.

Acknowledgements This work is supported by the National Natural Science Foundation of China (NO. 62376229) and Natural Science Foundation of Chongqing (NO. CSTB2023NSCQ-LZX0027).

Author Contributions **Weihua Xu:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation. **Jinlong Li:** Data curation, Methodology, Formal analysis, Software, Visualization, Writing - original draft, Writing - review & editing.

Declarations

Competing interests We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Publication ethics We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

Intellectual property We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

References

1. Hancer E, Xue B, Zhang M (2022) Fuzzy filter cost-sensitive feature selection with differential evolution. Knowl-Based Syst 241:108259
2. Huang P, Yang X (2022) Unsupervised feature selection via adaptive graph and dependency score. Patt Recognit 127:108622
3. Sang B, Chen H, Yang L, Li T, Xu W (2021) Incremental feature selection using a conditional entropy based on fuzzy dominance neighborhood rough sets. IEEE Trans Fuzzy Syst 30:1683–1697
4. Yang L, Qin K, Sang B, Fu C (2022) A novel incremental attribute reduction by using quantitative dominance-based neighborhood self-information. Knowl-Based Syst 261:110200
5. Bai H, Li D, Ge Y, Wang J, Cao F (2022) Spatial rough set-based geographical detectors for nominal target variables. Inf Sci 586:525–539
6. Zhang X, Hou J (2023) A novel rough set method based on adjustable-perspective dominance relations in intuitionistic fuzzy ordered decision tables. Int J Approx Reason 154:218–241

7. Jiang H, Hu BQ (2022) On (o, g)-fuzzy rough sets based on overlap and grouping functions over complete lattices. Int J Approx Reason 144:18–50
8. Xie J, Hu BQ, Jiang H (2022) A novel method to attribute reduction based on weighted neighborhood probabilistic rough sets. Int J Approx Reason 144:1–17
9. Shu W, Yan Z, Chen T, Yu J, Qian W (2022) Information granularity-based incremental feature selection for partially labeled hybrid data. Intell Data Anal 26:33–56
10. Yang X, Chen H, Li T, Wan J, Sang B (2021) Neighborhood rough sets with distance metric learning for feature selection. Knowl-Based Syst 224:107076
11. Wu S, Wang L, Ge S, Hao Z, Liu Y (2023) Neighborhood rough set with neighborhood equivalence relation for feature selection. Knowl Inf Syst, pp 1–27
12. Liu K, Li T, Yang X, Yang X, Liu D, Zhang P, Wang J (2021) Granular cabin: An efficient solution to neighborhood learning in big data. Inf Sci 583:189–201
13. Wan J, Chen H, Yuan Z, Li T, Yang X, Sang B (2021) A novel hybrid feature selection method considering feature interaction in neighborhood rough set. Knowl-Based Syst 227:107167
14. Liu K, Tsang EC, Song J, Yu H, Chen X, Yang X (2020) Neighborhood attribute reduction approach to partially labeled data. Granul Comput 5:239–250
15. Shu W, Yu J, Chen T, Qian W (2023) Neighbourhood discernibility degree-based semisupervised feature selection for partially labelled mixed-type data with granular ball. Appl Intell 53:22467–22487
16. Huang D, Zhang Q, Li Z (2023) Semi-supervised attribute reduction for partially labeled categorical data based on predicted label. Int J Approx Reason 154:242–261
17. Liu K, Yang X, Yu H, Fujita H, Chen X, Liu D (2020) Supervised information granulation strategy for attribute reduction. Int J Mach Learn Cybern, pp 1–15
18. Gao C, Zhou J, Miao D, Yue X, Wan J (2021) Granular-conditional-entropy-based attribute reduction for partially labeled data with proxy labels. Inf Sci 580:111–128
19. Pan Y, Xu W, Ran Q (2022) An incremental approach to feature selection using the weighted dominance-based neighborhood rough sets. Int J Mach Learn Cybern 14:1217–1233
20. Xu W, Yang Y (2023) Matrix-based feature selection approach using conditional entropy for ordered data set with time-evolving features. Knowl-Based Syst 279:110947
21. Yang Y, Chen D, Zhang X, Ji Z, Zhang Y (2022) Incremental feature selection by sample selection and feature-based accelerator. Appl Soft Comput 121:108800
22. Cai M, Lang G, Fujita H, Li Z, Yang T (2019) Incremental approaches to updating reducts under dynamic covering granularity. Knowl-Based Syst 172:130–140
23. Jiang Z, Liu K, Song J, Yang X, Li J, Qian Y (2021) Accelerator for crosswise computing reduct. Appl Soft Comput 98:106740
24. Liu K, Li T, Yang X, Chen H, Wang J, Deng Z (2023) Semifree: Semisupervised feature selection with fuzzy relevance and redundancy. IEEE Trans Fuzzy Syst 31:3384–3396
25. Zhang P, Li T, Yuan Z, Luo C, Liu K, Yang X (2022) Heterogeneous feature selection based on neighborhood combination entropy. IEEE Trans Neural Netw Learn Syst, pp 1–14
26. Xu W, Yuan K, Li W, Ding W (2022) An emerging fuzzy feature selection method using composite entropy-based uncertainty measure and data distribution. IEEE Trans Emerg Top Comput Intell 7:76–88
27. Liu Y, Zheng L, Xiu Y, Yin H, Zhao S, Wang X, Chen H, Li C (2020) Discernibility matrix based incremental feature selection on fused decision tables. Int J Approx Reason 118:1–26
28. Sheng K, Wang W, Xf Bian, Dong H, MA J (2020) Neighborhood discernibility degree incremental attribute reduction algorithm for mixed data. Acta Electonica Sin 48:682
29. Lin R, Li J, Chen D, Huang J, Chen Y (2021) Attribute reduction in fuzzy multi-covering decision systems via observational-consistency and fuzzy discernibility. J Intell Fuzzy Syst 40:5239–5253
30. Li X, Tang J, Hu B, Li Y (2022) Indiscernibility and discernibility relations attribute reduction with variable precision. Sci Prog 2022:1–11
31. Xia S, Liu Y, Ding X, Wang G, Yu H, Luo Y (2019) Granular ball computing classifiers for efficient, scalable and robust learning. Inf Sci 483:136–152
32. Xia S, Peng D, Meng D, Zhang C, Wang G, Giem E, Wei W, Chen Z (2022) Ball k-means: Fast adaptive clustering with no bounds. IEEE Trans Patt Anal Mach Intell 44:87–99
33. Xia S, Zhang H, Li W, Wang G, Giem E, Chen Z (2020) Gbnrs: A novel rough set algorithm for fast adaptive attribute reduction in classification. IEEE Trans Knowl Data Eng 34:1231–1242
34. Chen Y, Wang P, Yang X, Mi J, Liu D (2021) Granular ball guided selector for attribute reduction. Knowl-Based Syst 229:107326
35. Zhang P, Li T, Yuan Z, Luo C, Wang G, Liu J, Du S (2022) A data-level fusion model for unsupervised attribute selection in multi-source homogeneous data. Inf Fusion 80:87–103
36. Hu Q, Yu D, Liu J, Wu C (2008) Neighborhood rough set based heterogeneous feature subset selection. Inf sci 178:3577–3594
37. Yuan Z, Zhang X, Feng S (2018) Hybrid data-driven outlier detection based on neighborhood information entropy and its developmental measures. Expert Syst Appl 112:243–257
38. Dheeru D, Casey G (2019) Uci machine learning repository http://archive.ics.uci.edu/ml. irvine, ca: University of california. School Inf Comput Sci 25:27
39. Hu M, Tsang EC, Guo Y, Xu W (2021) Fast and robust attribute reduction based on the separability in fuzzy decision systems. IEEE Trans Cybern 52:5559–5572
40. Demšar J (2006) Statistical comparisons of classifiers over multiple data sets. J Mach Learn Res 7:1–30

Weihua Xu received the Ph.D. degree in mathematics from School of Sciences, Xian Jiaotong University, Xian, China, in 2007, and the M.Sc. degree in mathematics from School of Mathematics and Information Sciences, Guangxi University, Nanning, China, in2004.He is currently a Professor with the College of Artificial Intelligence, Southwest University, Chongqing, China. He has authored or coauthored 5 monographs and more than 240 articles in international journals. His current research interests include granular computing, cognitive computing, and information fusion. Dr. Xu also serves as a Senior Member of Chinese Association for Artificial Intelligence (CAAI).

Jinlong Li is currently pursuing the B.Sc. degree from the College of Artificial Intelligence, Southwest University, Chongqing, China. His current research interests include fuzzy sets, attribute reduction and granular computing.