

# Mejora de la clasificación de la diferenciación de la precisión de las variables en la gestión de la educación superior con inteligencia artificial

Chao Dong<sup>a</sup> y Yan Guo<sup>b,\*</sup>

<sup>a</sup>Universidad de Finanzas y Economía de Ningbo, Ningbo, China

<sup>b</sup>Universidad Tecnológica de Ningbo, Ningbo, China

**Abstracto** La amplia aplicación de la tecnología de inteligencia artificial en diversos campos ha acelerado el ritmo de exploración de la información oculta detrás de grandes cantidades de datos. La gente espera utilizar métodos de minería de datos para realizar investigaciones eficaces sobre la gestión de la educación superior, y el algoritmo de clasificación del árbol de decisiones como método de análisis de datos en la tecnología de minería de datos, la precisión de clasificación de alta precisión, los resultados de decisión intuitivos y la alta capacidad de generalización lo convierten en un método más ideal de gestión de la educación superior. Con el objetivo de mejorar la sensibilidad del procesamiento de datos y la clasificación del árbol de decisiones a los datos ruidosos, este documento propone mejoras correspondientes y propone un estándar de selección de atributos de conjunto aproximado de precisión variable basado en la función de escala, que considera tanto la precisión de aproximación ponderada como el valor del atributo. El número mejora la capacidad antiinterferente de los datos de ruido, reduce el sesgo en la selección de atributos y mejora la precisión de la clasificación. Al mismo tiempo, se introducen el umbral del factor de supresión, el soporte y la confianza en el proceso de poda previa del árbol, lo que simplifica la estructura del árbol. Los experimentos comparativos en conjuntos de datos estándar muestran que el algoritmo mejorado propuesto en este documento es mejor que otros algoritmos de árboles de decisión y puede realizar eficazmente la clasificación diferenciada de la gestión de la educación superior.

Palabras clave: Precisión variable, función de escala, gestión de la educación superior, árbol de decisión

## 1. Introducción

Con la aceleración de la popularización de la educación superior en nuestro país, la competencia nacional e internacional a la que se enfrenta la educación superior se está volviendo cada vez más feroz, y la supervivencia y el desarrollo de las universidades se enfrentan a grandes desafíos [1]. Los gerentes están pensando en un problema de este tipo: cómo hacer que la universidad obtenga la ventaja del desarrollo en la competencia y se convierta en el líder. El autor cree que el uso de la estrategia de diferenciación es una forma de que las universidades mejoren sus propias capacidades de desarrollo y mantengan sus propias ventajas competitivas.

La estrategia de diferenciación es una de las estrategias básicas de competencia de las empresas propuestas por el maestro de la estrategia Michael Porter. Esta teoría se aplica a la investigación de las organizaciones e instituciones educativas, y puede utilizarse para analizar la estrategia de desarrollo de las universidades [2-4]. Con la creciente diversificación de la educación superior, la competencia entre las universidades de nuestro país se ha vuelto cada vez más obvia, y se ha creado una situación en la que los fuertes se hacen más fuertes y los débiles se debilitan. Esto también demuestra que el modelo unilateral de gestión escolar de mil escuelas no ha podido satisfacer las necesidades de la economía de mercado, ni puede satisfacer el rápido desarrollo de la educación superior [5-7]. Las universidades deben reflexionar sobre su propia fuerza, posicionamiento y objetivos, buscar un espacio de desarrollo adecuado en función de sus

\*Autor correspondiente: Yan Guo, Ningbo Tech University, Ningbo 315100, China. Correo electrónico: guoyanbox@yeah.net.

Las escuelas de negocios deben desarrollar sus propias ventajas y características, fortalecer su propia rareza, reducir la amenaza de una competencia homogénea y formar una ventaja comparativa diferente de otras escuelas [8, 9]. Mejorar y desarrollar su propia competitividad básica y construir una ventaja competitiva única. Esta es la implementación de la llamada estrategia de diferenciación.

La diversificación de las universidades es el consenso formado por los círculos de la educación superior en mi país que entran en la etapa de popularización de la educación superior. Sin embargo, la situación real de la educación superior de mi país está lejos de las exigencias de diversidad y características. En la actualidad, hay dos estrategias que prevalecen entre las universidades: una es expandir rápidamente la escala, fusionarse, ampliar la matrícula, adquirir terrenos y construir casas, y utilizar todos los medios para crecer más; la otra es fortalecer la fuerza de la investigación científica e introducir doctores, profesores e incluso académicos con una alta inversión [10]. Recompensar los logros de la investigación científica, lo que se manifiesta de manera destacada en el aumento del nivel de funcionamiento de las escuelas. Esto ha provocado una grave convergencia en el desarrollo de diferentes universidades, lo que resulta en similitudes y funciones de servicio superpuestas. Esto ha llevado a la escalada continua de la competencia viciosa entre las escuelas, y se desperdician, abusan y dejan sin usar valiosos recursos educativos. La distancia entre la educación y la realidad social está aumentando [11]. En China, la construcción de infraestructura universitaria, equipos de talentos de información y recursos de información educativa ha logrado un gran progreso, pero en realidad se ha centrado en cambiar la estructura de enseñanza tradicional centrada en el profesor y construir un nuevo tipo de enseñanza combinada de asignaturas principales. Menos del 5% de las escuelas están integradas con el objetivo de la estructura. Este era un problema común en el proceso de informatización de la educación en ese momento, y también era un cuello de botella que restringía el desarrollo en profundidad de la informatización de la educación en mi país [12]. En este contexto, el diseño del sistema de indicadores se ajusta junto con la integración de la tecnología de la información y la enseñanza, discutiendo los puntos calientes y la urgencia de la demanda, y cambiando el enfoque de la investigación al proceso de aplicación de la información de la enseñanza.

Las características de las universidades surgen de sus diferencias. La raíz de las diferencias entre universidades radica en las diferencias en la demanda de productos de educación superior por parte de la sociedad y del mercado. La diferencia determina que las características y la individualización de la universidad sean la fuente de la ventaja competitiva sostenible de la universidad.

Por lo tanto, la universidad debe crear su propia competitividad única creando diferencias sobre la base de un estudio comparativo completo de los competidores y las necesidades sociales. Es decir, debemos hacer un uso completo de la estrategia de diferenciación para garantizar que la escuela tenga una capacidad sobresaliente para administrar una escuela y una excelente calidad de educación para tener la capacidad de competir continuamente. La gestión de la educación superior se muestra en la Figura 1.

## 2. Clasificación diferenciada de la gestión de la educación superior basada en inteligencia artificial de precisión variable

La disciplina es el elemento básico de las universidades, el centro de reunión de talentos y la base para que profesores y estudiantes desempeñen sus funciones [13]. Para que una escuela sea distintiva, centrada, de marca, competitiva y competitiva es inseparable de disciplinas y grupos de disciplinas únicos y ventajosos. El académico estadounidense Fiere [14] señaló: "No importa dónde, el trabajo de la educación superior se compone de dos modos básicos de cruce vertical y horizontal según las disciplinas e instituciones". En cierto sentido, la competitividad central de las universidades se concentra en sus disciplinas [15]. A nivel de construcción, las características de funcionamiento de las escuelas superiores y universidades se reflejan en última instancia en las características disciplinarias de la escuela. El indicador principal para juzgar si una universidad tiene competitividad central es su disciplina, y hay dos criterios para juzgar sus ventajas: uno es si está a la vanguardia del desarrollo científico y tecnológico del mundo; el otro es si se adapta a la tendencia del desarrollo social y económico [16]. Una universidad tiene su competitividad central solo cuando tiene una ventaja absoluta o relativa en una disciplina o grupo de disciplinas. La construcción disciplinaria, por un lado, debe esforzarse por fortalecer la construcción de disciplinas clave, utilizar disciplinas superiores para demostrar la radiación y promover el desarrollo coordinado de disciplinas relacionadas; por otro lado, también presta atención al desarrollo condensado de disciplinas y al cultivo de líderes disciplinarios. Existe un cierto grado de competencia entre los diversos departamentos dentro de la universidad debido al uso de los recursos limitados de la escuela. Es imposible que la escuela distribuya fondos y personal destacado de manera uniforme en toda la escuela [17, 18]. Por lo tanto, cuando las escuelas consideran la diferenciación, deben combinar la historia de la escuela.

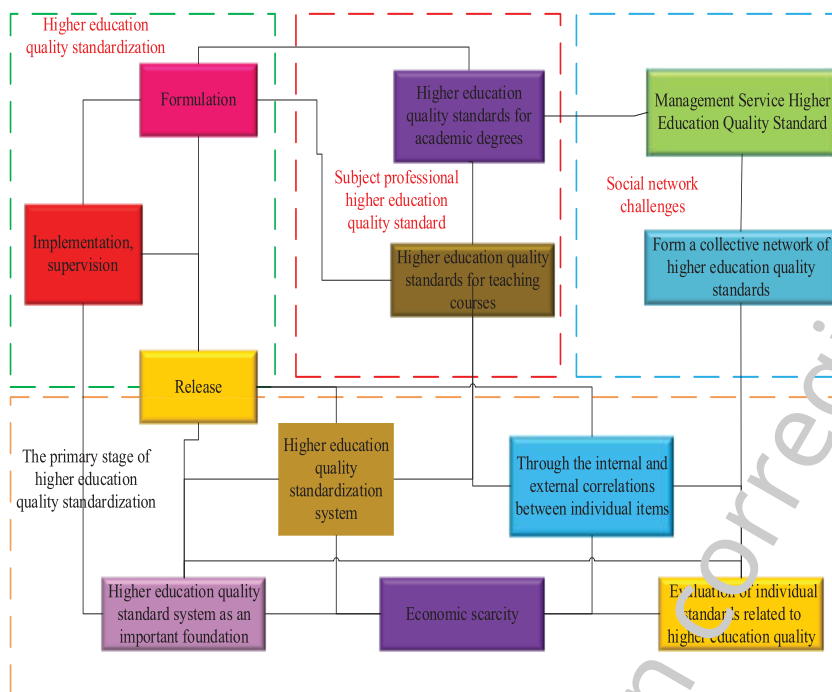


Fig. 1. Gestión de la educación superior.

Las condiciones, el ambiente cultural y otras condiciones realistas determinan las disciplinas clave de apoyo y formación de la escuela, porque estas disciplinas son los recursos básicos adquiridos por la escuela en la acumulación histórica a largo plazo. Es difícil que se establezcan otras escuelas mediante una simple duplicación o imitación en un corto período de tiempo. Solo cuando la escuela invierta fondos y personal en esta dirección, las disciplinas relacionadas pueden convertirse en disciplinas líderes en el país e incluso en el mundo.

### 2.1. Algoritmo de clasificación de árboles de decisión de precisión variable basado en una función de escala

El árbol de decisión es un método simple pero ampliamente utilizado en modelos de clasificación y predicción. El árbol de decisión divide un gran conjunto de datos en partes más pequeñas [19–21]. Esta parte más pequeña del conjunto de datos se denomina nodo terminal. A cada nodo terminal se le asigna una etiqueta de clase como marca. Los nodos no terminales están compuestos por nodos raíz y otros nodos internos, e incluyen condiciones de prueba de atributos para distinguir registros de datos con diferentes características. Cuando el subconjunto de datos que se ha particionado no se puede particionar aún más de acuerdo con los criterios de partición definidos previamente, se termina todo el proceso de partición.

En esta ocasión, obtenemos un árbol de decisiones con capacidades de partición de la toma de decisiones. Los árboles de decisiones se utilizan en muchos campos. Por ejemplo, en el marketing de bases de datos, los árboles de decisiones se pueden utilizar para agrupar diferentes grupos de clientes y ayudar a los especialistas en marketing a realizar promociones correspondientes para diferentes niveles de grupos de clientes para lograr una mayor tasa de respuesta. Lograr las ventas de productos esperadas.

El algoritmo de clasificación del árbol de decisión finalmente genera un árbol [22]. Al observar el árbol, podemos entender fácilmente la estrategia de clasificación correspondiente. Sin embargo, los datos reales utilizados por Shan Yu son demasiado grandes y el árbol que obtenemos es demasiado complejo. La construcción de una estructura de árbol con una estructura simple y un buen efecto de clasificación siempre ha sido el foco de atención de los investigadores. En los últimos años, con el auge del fenómeno del "calor rugoso", la teoría de conjuntos rugosos y de conjuntos rugosos de precisión variable también se ha unido al ejército de algoritmos de clasificación [21]. La práctica ha demostrado que la fusión de la teoría de conjuntos rugosos y la teoría del árbol de decisión puede construir una estructura de árbol de decisión más optimizada. Basándose en estudios previos, este artículo introduce el concepto de función de escala y teoría de precisión variable en el algoritmo de clasificación del árbol de decisión, y utiliza inhibidores para limitar el proceso de crecimiento del árbol durante el proceso de construcción del árbol, y propone una nueva clasificación de árbol de decisión mejorada.

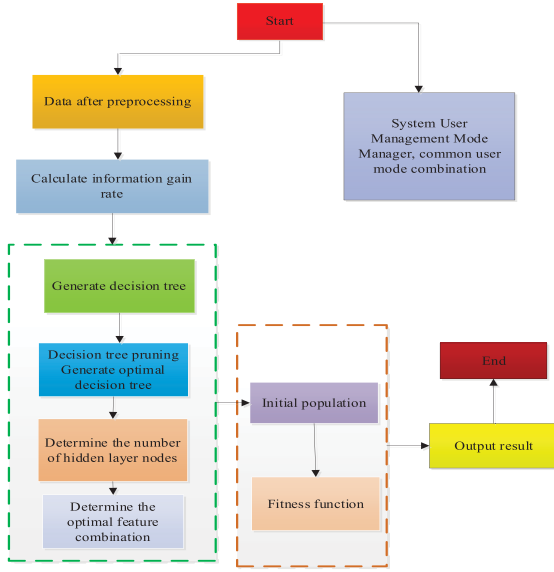


Fig. 2. Algoritmo de clasificación de árboles de decisión.

Algoritmo de clasificación del árbol de decisión. El algoritmo de clasificación del árbol de decisión se muestra en la figura 2.

## 2.2 Teoría de precisión variable

La teoría de conjuntos aproximados fue propuesta por el científico polaco Pawlak en 1982. Esta teoría es una nueva herramienta matemática para tratar con el conocimiento vago e impreciso [24]. La teoría de precisión variable se basa en la extensión de la teoría de conjuntos aproximados. La mayor diferencia con la teoría de conjuntos aproximados es que permite un cierto error de clasificación al dividir, en lugar de pertenencia y no pertenencia absolutas. La teoría de conjuntos aproximados de precisión variable optimiza la teoría de conjuntos aproximados mediante la limitación de la precisión de la clasificación. El modelo de conjunto aproximado de precisión variable mejora la capacidad antiinterferencia de los datos de ruido mediante la introducción de parámetros, pero tiene un mayor impacto en la selección de valores de precisión de las variables. Para seleccionar el valor de precisión de la variable adecuado de manera específica para cumplir con los diferentes requisitos de reglas de los diferentes usuarios, la variable de escala  $R(s)$  se introducen para establecer un modelo de decisión dinámico.

Supongamos que la variable de escalas  $s=1,2, \dots, N$ , la función de escala  $R(s) \in [0.5, 1]$ , la precisión de la división  $\alpha = 0.10$ , entonces  $R(1) = 0.5, R(2) = 0.6, R(3) = 0.7, R(4) = 0.8, R(5) = 0.9$ . Cuanto más precisa sea la división de la escala, más detallada será la expresión del conocimiento.

## Multiescala

toma de decisiones

sistema

$D = (U, A, V, R(s)), A = CD$ ,  $C$  es el conjunto de atributos condicionales,  $D$  es el conjunto de atributos de decisión y el subconjunto  $B \subset D$  del universo  $U$  se divide en:  $Subtítulo = \{m_1, m_2, \dots, m_{norte}\}, R(s) \in [0.5, 1]$  son variables de escala.  $\forall Y \in Subtítulo$ , las aproximaciones superior e inferior de  $R(s)$ ,  $Y$  relativa a  $R(s)$  son:

$$B^*(R(s)) = \bigcup \{mi | d\alpha(S, M) \leq 1 - R(s)\} \quad (1)$$

$$B_*(R(s)) = \bigcup \{mi | d\alpha(S, M) \geq R(s)\} \quad (2)$$

Entre ellos,  $d\alpha(Y/X) = 1 - \frac{|in \cap out|}{|in \cup out|}$  representa la tasa de clasificación errónea de clasificar  $Y$  como conjunto  $X$ .

## 2.3. Algoritmo mejorado de árbol de decisión de precisión variable basado en función de escala

La diferencia entre la clasificación del árbol de decisión basada en un conjunto aproximado de precisión variable y la clasificación del árbol de decisión según la teoría tradicional de conjuntos aproximados es que los criterios para la selección de atributos no se limitan a seleccionar los atributos con el área libre más grande, el área no libre más pequeña o el área límite más pequeña como atributos divididos, la precisión promedio ponderada también se propone como un nuevo criterio de selección de atributos [25]. La introducción de la precisión de clasificación (valor de precisión variable) permite que el árbol de decisión se puede durante el proceso de crecimiento, lo que reduce la complejidad del árbol de expansión; al mismo tiempo, también amplía el rango del área libre, puede tolerar la existencia de algunos datos de ruido y mejora la robustez del árbol. Sin embargo, el árbol de decisión bajo el modo de conjunto aproximado de precisión variable aún tiene el sesgo de selección de atributos divididos hacia aquellos atributos con más valores de atributo [26]. Dado que las formas de los atributos y los campos de adquisición de datos son diferentes, el atributo con el mayor número de valores de atributo no es necesariamente el atributo con el mejor efecto de clasificación, por lo que el sesgo de atributo siempre ha sido un factor importante que afecta el efecto de clasificación.

Por un lado, en el sistema de índices de evaluación de la informatización universitaria actualmente establecido en nuestro país, éste se compone mayoritariamente de indicadores de dos niveles. Es difícil realizar una investigación profunda y detallada sobre los indicadores de cada nivel [27]. La relación difusa de los niveles de evaluación hace que algunos indicadores tengan una mayor correlación lo que facilita que los mismos factores ocupen una gran proporción al calcular el peso del índice, lo que hace que la

El resultado de la evaluación no es razonable; por otra parte, debido a la relación difusa en el sistema de índice de evaluación compuesto por indicadores de dos niveles, la mayoría de los indicadores tienen características de integración que hacen que la incorporación de algunos indicadores de actividad de información importantes y la medición que incluye el método de promedio aritmético carezca de contenido de análisis jerárquico profundo. En un sistema de múltiples índices, a menudo hay una superposición de información grave o multicolinealidad de variables explicativas entre los indicadores, lo que provocará imprecisiones estadísticas. Los resultados son difíciles de distinguir la contribución de diferentes parámetros o factores. Además, el resultado también es un valor de comparación relativo bajo la premisa de determinar un cierto punto de referencia, que solo tiene un significado relativo, no un significado absoluto.

En este artículo, que aborda las desventajas de los métodos de minería de árboles de decisión existentes, como la dificultad en la selección de atributos, la débil capacidad antirruído y la incapacidad de satisfacer las necesidades específicas de los usuarios, se combina el conocimiento de la función de escala y el conjunto aproximado de precisión variable, y se propone un algoritmo teórico mejorado (MSVPRSDT) que combina la función de escala y el conjunto aproximado de precisión variable. En el algoritmo, se proporciona un nuevo criterio de selección de atributos y se vuelven a proporcionar la condición de división de parada de nodo y el método de predicción de la etiqueta de clase [28].

En el proceso de construcción del árbol, considerando la capacidad anti-ruido del árbol, el sesgo de los atributos y las necesidades específicas de usuarios específicos, se propuso una nueva división de atributos de importancia estándar-atributo, que consideró de manera integral la función de escala y la aproximación ponderada. Precisión de clasificación y número de valores de atributo.

El método de predicción de la etiqueta de clase del clasificador de árbol de decisión tradicional se basa en el nodo raíz como punto de partida y el árbol de decisión se recorre capa por capa a lo largo de la rama de prueba con el mismo valor de atributo hasta llegar al nodo hoja. En este momento, la etiqueta de clase del nodo hoja es el resultado de la predicción de la etiqueta de clase. Sin embargo, debido a la existencia de datos de ruido y algunos valores predeterminados, el efecto de clasificación de este método no es muy bueno. Por lo tanto, este capítulo utiliza el método de predicción de clasificación basado en el grado de coincidencia.

La ruta entre el nodo raíz y cada nodo hoja en el árbol de decisión se define como una regla de decisión. Cada atributo de condición de prueba en la regla es el antecedente de la regla, y la etiqueta de clase de un subnodo se denomina consecuente de la regla.

Suponiendo que la muestra de prueba es  $S$  y la regla de decisión es  $r$ , entonces la muestra, el grado de coincidencia con

La regla de decisión  $r$  y la tasa de precisión de la predicción son:

$$fósford(muestra, regla) = \frac{|condición_{muestra}^{regla}|}{|condición_{regla}|} \quad (3)$$

$$exactitud = \frac{yo}{\frac{elementos\ correctos}{yOelementos\ totales}} \quad (4)$$

Dónde  $|condición_{muestra}^{regla}|$  representa el número de muestras de prueba y condiciones cumplidas en regla,  $|condición_{regla}|$  representa el número total de condiciones en la regla,  $yOelementos\ totales$  es el número de elementos en el conjunto de datos de predicción,  $yo$  es el número de etiquetas de clase correctas [20].

De la definición se desprende que:  $fósford(muestra, regla) \in [0, 1]$ . Cuando se realiza la predicción de la etiqueta de clase de la muestra de prueba, se selecciona como resultado de la predicción la etiqueta de clase de la regla con el mayor grado de coincidencia. Cuando hay varias reglas con el mismo grado de coincidencia, se selecciona como resultado de la predicción la etiqueta de categoría de la regla que aparece con mayor frecuencia.

Criterios para la selección de atributos divididos: este documento utiliza un algoritmo de clasificación de árbol de decisión mejorado que tiene en cuenta la función de escala, la teoría de conjuntos aproximados y el número de valores de atributo. Diferentes funciones de escala pueden satisfacer los diferentes requisitos de precisión de toma de decisiones de diferentes usuarios. Sobre la base de la función de escala, la aplicación de la teoría de conjuntos aproximados para tratar el conocimiento difuso e inexacto puede mejorar eficazmente la robustez del algoritmo [29]. El algoritmo de clasificación de árbol de decisión tradicional tiene las características de seleccionar atributos con una gran cantidad de valores, y la experiencia real nos dice que los atributos con una gran cantidad de valores de atributo no son necesariamente suficientes para mejorar el efecto de clasificación. Para deshacerse de este inconveniente, se consideran los atributos de la cantidad de valores en el efecto de clasificación. Este documento utiliza la precisión de clasificación aproximada ponderada de la función de escala y la suma del número de valores de atributo para seleccionar atributos en la importancia de atributo más estándar del grado de selección de atributo dividido, e introduce factores de supresión y niveles de confianza para cortar el árbol de decisión durante el proceso de construcción del árbol. En comparación con el árbol construido por el algoritmo tradicional, el árbol construido tiene una estructura simple, mayor precisión de clasificación y se ha mejorado aún más la velocidad de ejecución.

### 3. Análisis de experimentos y resultados

#### 3.1 Análisis de casos

Los ajustes de parámetros en el algoritmo MSVPRSDT son los siguientes:  $R(s) = 0.8$ ,  $\alpha = 0.5$ , umbral del factor de supresión  $\text{del}_{\min} = 0.8$ . Utilice la fórmula para calcular la importancia de cada atributo para obtener:

$$\text{Firma}(\text{Aplicación de información para la enseñanza}) = 0.495, \quad (5)$$

$$\text{Firma}(\text{Información del método de enseñanza}) = 0.315, \quad (6)$$

$$\text{Firma}(\text{automatización de oficinas}) = 0.755 \quad (7)$$

$$\text{Firma}(\text{Gestión de la información}) = 0.512 \quad (8)$$

Por lo tanto, se selecciona la aplicación de informatización de la enseñanza como el nodo de división actual.

El atributo de aplicación de la enseñanza incluye dos atributos (alto, normal) y los valores subconjuntos que contienen.

son  $t_{\text{alto}} = \{1, 2, 3, 4, 8, 12, 14\}$ ,  $t_{\text{normal}} = \{5, 6, 7, 9, 10, 11, 13\}$ . La clase Las etiquetas de estos dos subconjuntos son inconsistentes, para el subconjunto  $t_{\text{normal}}$ , la confianza del subconjunto es  $\text{conferencia}(\text{normal}) = \text{GRAMO}/7 > 0.8$ , por lo que el nodo deja de dividirse y se marca como un nodo hoja [23]. La etiqueta de clase del nodo hoja utiliza la etiqueta de clase más grande en el reconocimiento de subconjuntos.  $t_{\text{normal}}$  Por lo tanto, se obtiene una rama con una longitud de arco normal y un nodo de hoja marcado como sí. Para otro subconjunto  $t_{\text{alto}}$ , la confianza del nodo correspondiente es menor que el umbral de confianza dado 0.8, por lo que se vuelve al primer paso para este subconjunto y se continúa con la ejecución. Después del cálculo,  $\text{Sig}(\text{información del método de enseñanza}) = 0.626 < 0.8$ , se pasa al último paso para continuar con la ejecución y se encuentra mediante la observación. Hay tres valores de atributo bajo el atributo de informatización de los métodos de enseñanza (construcción de infraestructura, construcción del equipo de talentos de informatización y construcción de recursos de información), entre los cuales la confianza del valor de atributo construcción de infraestructura y construcción del equipo de talentos de informatización es  $\text{conf}(\text{construcción del equipo de talentos de informatización}) = 1 > 0.8$ ,  $\text{conf}(\text{construcción de infraestructura}) = 1 > 0.3$ , el nivel de confianza del valor de atributo construcción de recursos de información  $\text{conf}(\text{construcción de recursos de información}) = 0.5 < 0.8$ , por lo que se obtiene una longitud de arco para la construcción de infraestructura y

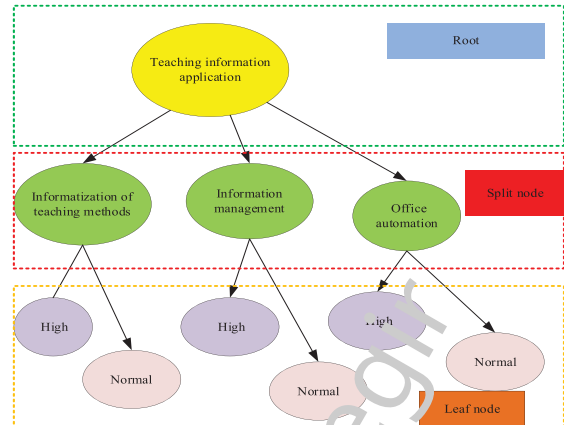


Fig. 3. Árbol de decisión generado por el algoritmo MSVPRSDT.

Nodos de hoja Una rama marcada como no y otra rama con una longitud de arco para la formación de equipos de talentos de tecnología de la información y un nodo de hoja marcado como sí. Hasta ahora, se considera que todos los subconjuntos se han ejecutado y la estructura del árbol de decisión final se muestra en la figura 3 [30].

Es  $R(s) = 0.9$  y  $T(s) = 1$  se obtiene el mismo árbol de decisión, la complejidad es 8, el número de hojas es 5, la profundidad del árbol es 3 y la precisión de la regla de decisión es del 90%. El árbol de decisión con  $R(s) = 0$ . Se selecciona 7, la complejidad es 10, el número de hojas es 8, la profundidad del árbol es 4, la precisión de la regla es del 70%.  $R(s) = 0.9$ ,  $R(s) = 1$  tienen la misma estructura de árbol. La función de escala  $R(s) = 0.9$ , en comparación con  $R(s) = 0.7$ , es mejor que  $R(s) = 0.7$  y  $R(s) = 0.9$  en términos de precisión y estructura de árbol. Por lo tanto, para este ejemplo, la mejor función de escala es  $R(s) = 0.9$ .

#### 3.2 Análisis de los resultados experimentales sobre el conjunto de datos estándar

Hay cuatro parámetros variables en el algoritmo MSVPRSDT: función de escala  $R(s)$ , umbral de confianza  $\text{minconf}$ , coeficiente de ajuste  $\alpha$ , factor de supresión  $\text{del}_{\min}$ . Se seleccionaron tres conjuntos de datos para el experimento. Las figuras 4 y 5 muestran respectivamente la precisión de clasificación de diferentes funciones de escala.  $R(s)$  en un solo conjunto de datos y el tamaño del árbol de expansión.

La figura 4 muestra que cuando la función de escala está entre 0.8 y 0.9, la precisión de clasificación del algoritmo MSVPRSDT en estos dos conjuntos de datos es relativamente alta, y la precisión de clasificación de la función de escala 0.88 de estos tres conjuntos de datos permanece básicamente sin cambios. La figura 5 muestra que la escala de

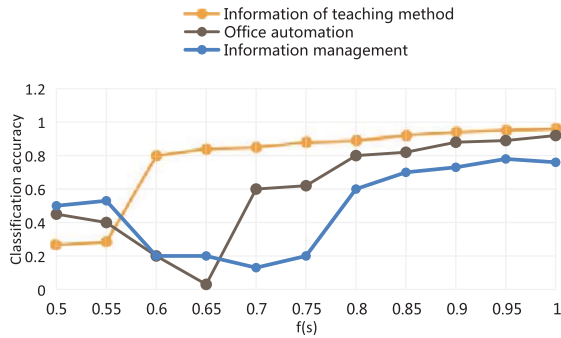


Fig. 4. Comparación de la precisión de la clasificación bajo diferentes funciones de escala  $f(s)$ .

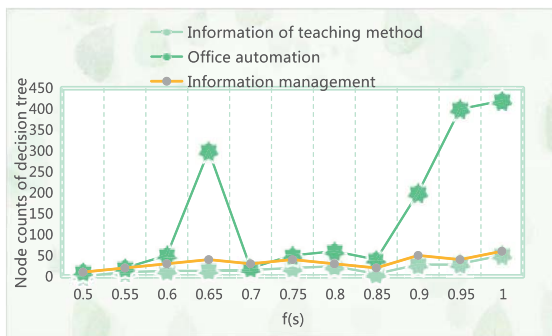


Fig. 5. Comparación del tamaño del árbol bajo diferentes funciones de escala  $f(s)$ .

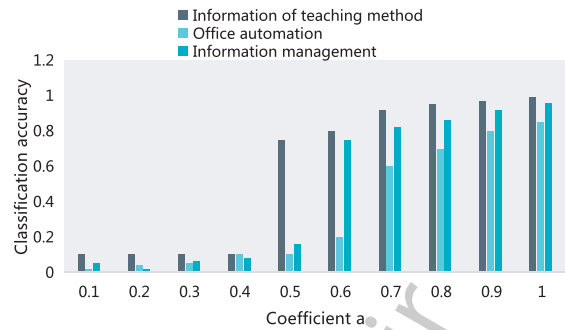


Fig. 6. Comparación de la precisión de la clasificación bajo diferentes coeficientes.

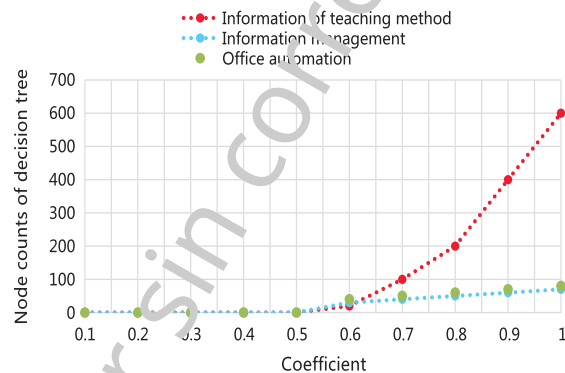


Fig. 7. Comparación del tamaño de los árboles bajo diferentes coeficientes.

El árbol de expansión es proporcional a la función de escala. Cuando mayor sea  $f(s)$ , cuanto mayor sea la escala del árbol. La escala del árbol de expansión  $f(s)=1$  aumenta repentinamente, lo que indica que  $f(s)$  La función tiene un mejor efecto de clasificación cuando está cerca de 1, y  $f(s)$  No puede estar demasiado cerca de 1.

Las figuras 6 y 7 muestran respectivamente la precisión de clasificación de diferentes coeficientes  $a$  en el conjunto de datos y el tamaño del árbol. La figura 6 muestra que cuando el coeficiente  $a < 0.5$ , el cambio del coeficiente  $a$  tiene poco efecto en la precisión de la clasificación y la escala del árbol de expansión. Cuando  $a > 0.5$ , a medida que aumenta el coeficiente  $a$ , la precisión de la clasificación y la escala del árbol de expansión del algoritmo MSVPRSDT en el conjunto de datos tienen una tendencia a volverse más pequeñas. Cuando el coeficiente  $a$  aumenta, la precisión de la clasificación y la escala del árbol de expansión del algoritmo MSVPRSDT en el conjunto de datos tienen una tendencia a volverse más pequeñas.  $a/f=0.5$ . Tanto la precisión de la clasificación como la escala del árbol de expansión tienen los mejores resultados.

El algoritmo de árbol de decisión de precisión variable basado en la función de escala es el método de análisis de datos más utilizado en la tecnología de clasificación. Utiliza un conjunto de muestras de entrenamiento como datos de entrada e implementa la descripción de cada clase de decisión.

mediante la aproximación de la función objetivo discreta. Descubrir la función resultante en forma de árbol. Luego clasificar la nueva muestra cuya clase pertenece a la muestra de prueba a través de la función obtenida por inducción, predecir a qué clase de decisión pertenece la nueva muestra y calcular la precisión de la clasificación para evaluar el desempeño de la función aprendida. El algoritmo del árbol de decisión consiste en descubrir las reglas de clasificación ocultas detrás de una gran cantidad de datos en forma de un árbol intuitivo.

De la figura 8 se desprende que cuanto mayor sea el umbral de confianza  $minconf$ , cuanto mayor sea la tasa de preparación de clasificación del algoritmo, pero la expresión de la página de escala del árbol de expansión, cuando el umbral de confianza  $minconf$  se obtiene un valor apropiado, se puede obtener una mayor tasa de precisión de clasificación y un árbol de expansión más pequeño en la misma escala de tiempo.

La figura 9 muestra que el cambio del factor de supresión es inversamente proporcional a la tasa de precisión de la clasificación y la escala del árbol de expansión.  $del_{min} > 0.9$ . La tasa de precisión de la clasificación comienza a disminuir y la escala del árbol de expansión también comienza a disminuir.



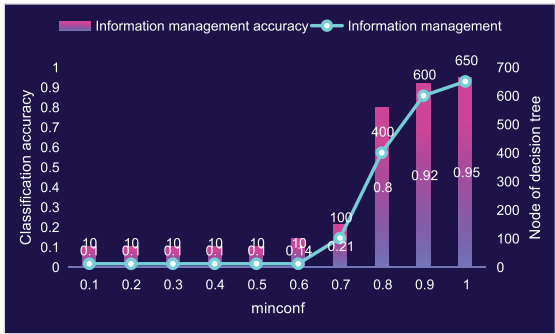


Fig. 8. Influencia del umbral de confianza en el algoritmo.

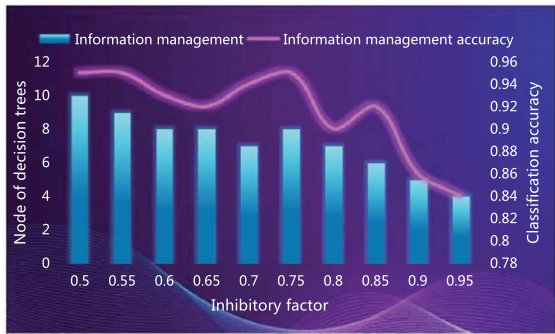


Fig. 9. Influencia del coeficiente a en el algoritmo.

Tabla 1

Resultados de la clasificación diferenciada mediante validación cruzada

Iteraciones	Número de decisiones sobre muestras de prueba		
	De hecho coherente con el muestra	De hecho inconsistente con el muestra	No puede ser clasificado
3	12	3	1
6	14	2	0
9	15	1	0
12	15	1	0
15	16	0	0
18	16	0	0

Basado en el algoritmo de clasificación de árboles de decisión de conjunto aproximado de precisión variable, este experimento propone un algoritmo de clasificación de árboles de decisión mejorado que utiliza una función de escala en el proceso de selección de atributos e introduce un factor de supresión en el método de poda previa de árboles para mejorar la generación de árboles de clasificación. La tasa de precisión es el propósito de reducir la complejidad del árbol de clasificación. Al final, se utilizan 16 conjuntos de datos estándar para el análisis experimental comparativo con otros algoritmos de clasificación de árboles de decisión que tienen mejores

Efectos de la clasificación. Finalmente, se analizan los cuatro parámetros en el proceso de ejecución del algoritmo.

La precisión de clasificación promedio del conjunto de prueba es del 77,5% y la cobertura de clasificación promedio es del 90%. Esto se encuentra dentro de un rango aceptable en este artículo. Un conjunto de reglas de decisión demasiado simplificado dará como resultado la incapacidad de clasificar o cubrir con precisión algunas muestras de productos nuevos y aplicaciones prácticas. En el proceso, los parámetros de ganancia de información se pueden ajustar de manera flexible en función de las características de los datos para obtener un conocimiento perceptivo de una escala razonable.

4. Conclusión

Algoritmo de árbol de decisión de precisión variable basado en la función de escala. Este artículo estudia las ventajas de los conjuntos aproximados, la teoría de conjuntos aproximados de precisión variable y la teoría de conjuntos aproximados de precisión variable en el campo de la clasificación, y propone un método que combina la función de escala y la teoría de conjuntos aproximados para clasificar conjuntos de datos. En los criterios de selección de atributos divididos, se consideran tanto la precisión de clasificación aproximada ponderable como el número de valores de atributo; en el método de pre-poda de árboles de decisión, se propone el concepto de factor de supresión para hacer que la estructura del árbol sea más simple e intuitiva. En comparación con los métodos tradicionales de clasificación de árboles de decisión, el algoritmo mejorado propuesto en este artículo tiene mejores resultados en la precisión de la clasificación y la complejidad del árbol de expansión. El algoritmo de clasificación mejorado del árbol de decisión propuesto en este artículo se aplica a la base de datos de gestión de la educación superior, y se diseña y realiza el submódulo de gestión de la educación superior: la clasificación del árbol de decisión, y se descubren los factores que influyen en la gestión de la educación superior, lo que tiene una gran influencia en la gestión de la educación superior. Aunque este artículo ha realizado ciertos resultados de investigación sobre el llenado de datos faltantes y la teoría de clasificación del árbol de decisión y la mejora del algoritmo. Sin embargo, debido a la complejidad de los datos reales, la amplitud del contenido de la investigación y la capacidad personal, todavía existen muchas deficiencias en el trabajo de investigación y aún quedan muchos problemas que deben estudiarse más a fondo y mejorarse en consecuencia. El método de poda en el algoritmo de árbol de decisión mejorado en este documento es diferente de la estrategia tradicional que solo utiliza el apoyo y la confianza. Aunque la precisión de la clasificación final mejora significativamente, la complejidad del árbol no se reduce significativamente. Por lo tanto, el algoritmo se ejecuta. Es mejor configurar



los parámetros del proceso de forma específica, o utilizar otros métodos de poda mejorados para reducir la complejidad del árbol de expansión.

## Expresiones de gratitud

Este trabajo cuenta con el apoyo del Proyecto de Programación de Ciencias Educativas de la Provincia de Zhejiang (2020SCG076).

## Referencias

- [1] B. Newman y ET Liu, Perspectiva sobre BRCA1, *Enfermedad de la mama* **10**(1998), 3–10.
- [2] H. Liang, BY Tsui, H. Ni, et al., Evaluación y diagnósticos precisos de enfermedades pediátricas utilizando inteligencia artificial[J], *Medicina natura* **25**(3) (2019), 433–438.
- [3] Z. Lv, Y. Han, AK Singh, G. Manogaran y H. Lv, Confiabilidad en sistemas IoT industriales basados en inteligencia artificial, *Transacciones IEEE sobre informática industrial* **17**(2) (2020), 1496–1504.
- [4] SA Aljawarneh, Revisión y exploración de herramientas innovadoras de aprendizaje ubicuo en la educación superior[J], *Revista de informática en la educación superior* **32**(1) (2020), 57–73.
- [5] F. Farivar, MS Haghighi, A. Jolfaei y M. Alazab, Inteligencia artificial para la detección, estimación y compensación de ataques maliciosos en sistemas ciberfísicos no lineales e IoT industrial, *Transacciones IEEE sobre informática industrial* **16**(4) (2018), 2716–2725.
- [6] MJ Steinhausen, PD Wagner, B. Narasimhan, et al., Combinación de datos de Sentinel-1 y Sentinel-2 para mejorar el mapeo del uso y la cobertura del suelo en las regiones monzónicas[J], *Revista internacional de observación de la Tierra y geoinformación aplicada* **73**(2018), 595–604.
- [7] R. Sundarasekar, PM Shakeel, S. Baskar, S. Kadry y G. Mastorakis, Calidad de servicio adaptativa consciente de la energía para una transferencia de datos confiable en redes de sensores acústicos submarinos, *Acceso IEEE* (2020), 80093–80103.
- [8] K. Colchester, H. Hagras, D. Alghazzawi, et al., Una encuesta sobre técnicas de inteligencia artificial empleadas para sistemas educativos adaptativos dentro de plataformas de aprendizaje electrónico[J], *Revista de investigación en inteligencia artificial y computación blanda* **7**(1) (2017), 47–64.
- [9] J. Chen, L. Ramanathan y M. Alazab, Modelado inteligente artificial integrado de big data holístico para mejorar la privacidad y la seguridad en la gestión de datos de ciudades inteligentes, *Microprocesadores y microsistemas* **81**103722.
- [10] Y. Chen, J. Xiong, W. Xu, et al., Un nuevo algoritmo de aprendizaje incremental y decremental en línea basado en una máquina de vectores de soporte variable[J], *Computación en clúster* **22**(3) (2019), 7435–7445.
- [11] G. Manogaran, S. Baskar, PM Shakeel, N. Chakramurti y R. Kumar, Análisis de video de vigilancia en tiempo real utilizando verificación de cuadro regresivo acelerado de transformada de dos bits, *Herramientas y aplicaciones multimedia* (2020), 1–18.
- [12] M. Alam, D. Le, JI Lim, et al., Clasificación de retinopatías mediante inteligencia artificial multitarea basada en aprendizaje automático supervisado[J], *Revista de medicina clínica* **8**(6) (2019), 872.
- [13] K. Okamoto, S. Kitamura, T. Kimura, et al., Características clinicopatológicas de los pólipos serrados como precursores del cáncer colorrectal: estado actual y tratamiento[J], *Revista de Gastroenterología y Hepatología* **32**(2) (2017), 358–367.
- [14] N. Padhy, RP Singh y SC Satapathy, Modelo de predicción de reutilización rentable y resistente a fallas mediante el uso de una red neuronal basada en un algoritmo genético adaptativo para aplicaciones web de servicio[J], *Computación en clúster* **22**(6) (2019), 14559–14581.
- [15] U. Fiore, A. De Santis, F. Perla, et al., Uso de redes generativas antagonicas para mejorar la eficacia de la clasificación en la detección de fraudes con tarjetas de crédito[J], *Ciencias de la información* **47**(9) (2019), 448–455.
- [16] HM Vo, C. Zhu, NA Diep, El efecto del aprendizaje combinado en el desempeño de los estudiantes a nivel de curso en la educación superior: un metaanálisis[J], *Estudios en Evaluación Educativa* **5**(3) (2017), 17–28.
- [17] M. Agaoglu, Predicción del desempeño del instructor utilizando técnicas de minería de datos en la educación superior[J], *Acceso IEEE* **1**(4) (2016), 2379–2387.
- [18] YJ Yang y CS Bang, Aplicación de inteligencia artificial en gastroenterología[J], *Revista Mundial de Gastroenterología* **25**(14) (2019), 16–19.
- [19] H. Xiao, B. Muthu y SN Kadir, Corrección de: Inteligencia artificial con robótica para la industria manufacturera avanzada utilizando un modelo de programación de números enteros mixtos asistido por robot, *Intelligent Service Robotics*. (2020). doi:10.1007/s11370-020-00330-x
- [20] TN Nguyen, B. Liu, NP Nguyen y J. Chou, Ciberseguridad de la red inteligente: ataques y defensas, *CPI 2020- Conferencia internacional sobre comunicaciones del IEEE 2020 (ICC)*, Dublín, Irlanda, 2020, págs. 1–6, doi: 10.1109/ICC40777.2020.9148850.
- [21] C. Kirttan, Wong, AS Bombback, U. Baber, et al., Dirección futura para el uso de inteligencia artificial para predecir y controlar la hipertensión[J], *Informes actuales sobre hipertensión* **20**(9) (2018), 75–77.
- [22] L. Patrício y R. Rieder, Visión artificial e inteligencia artificial en agricultura de precisión para cultivos de granos: una revisión sistemática[J], *Computadoras y electrónica en la agricultura* **15**(3) (2018), 69–81.
- [23] V. Cantisani, E. David, H. Grazhdani, et al., Evaluación prospectiva de la relación de deformación semicuantitativa y la elastografía de ondas de corte por ultrasonidos 2D cuantitativa (SWE) en asociación con la clasificación TIRADS para la caracterización de nódulos tiroideos[J], *Ultraschall in der Medizin-Revista Europea de Ultrasonido* **40**(4) (2019), 495–503.
- [24] M. Bagheri, A. Akbari y SA Mirbagheri, Control avanzado de la suciedad de las membranas en sistemas de filtración mediante técnicas de inteligencia artificial y aprendizaje automático: Una revisión crítica[J], *Seguridad de procesos y protección del medio ambiente* **12**(3) (2019), 229–252.
- [25] CC Wu, WD Hsu, MM Islam, et al., Un enfoque de inteligencia artificial para predecir de forma temprana el infarto de miocardio sin elevación del segmento ST en pacientes con dolor torácico[J], *Métodos y programas informáticos en biomedicina* **17**(3) (2019), 109–117.
- [26] AJ Lowe, MF Breed, H. Caron, et al., Correlaciones estandarizadas de diversidad genética e historia de vida para una mejor gestión de los recursos genéticos de los árboles neotropicales[J], *Diversidad y distribuciones* **24**(6) (2018), 730–741.
- [27] MT Duong, AM Rauschecker, JD Rudie, et al., Inteligencia artificial para educación de precisión en radiología[J], *Revista británica de radiología* **92**(3) (2019), 38–39.

- 652 [28] S. Zubaidah, NM Fuad, S. Mahanal, et al., Mejora de las habilidades de  
653 pensamiento creativo de los estudiantes a través de la investigación  
654 científica diferenciada integrada con mapas mentales[]], *Revista de*  
655 *educación científica turca***14**(4) (2017), 77-91.
- 656 [29] MF Byrne, N. Chapados, F. Soudan, et al., Diferenciación en tiempo real de  
657 pólipos colorrectales diminutos adenomatosos e hiperplásicos durante  
658 el análisis de videos inalterados de colonoscopia estándar utilizando un  
659 modelo de aprendizaje profundo[]],  
660 *Intestino***68**(1) (2019), 94-100.
- [30] R. Cerezo, M. Sànchez-Santillàn, MP Paule-Ruiz, et al., Patrones de  
interacción de los estudiantes en el LMS y su relación con el  
rendimiento: un estudio de caso en la educación superior[]],  
*Computadoras y educación***9**(6) (2016), 42-54.
- 661  
662  
663  
664

Prueba de autor sin corregir