

1 Mejora de la clasificación de la diferenciación de la 2 precisión de las variables en la gestión de la educación 3 superior con inteligencia artificial

4 Chao Dong y Yan Guob,*

5 *aUniversidad de Finanzas y Economía de Ningbo, Ningbo, China*

6 *bUniversidad Tecnológica de Ningbo, Ningbo, China*

7 **Abstracto** La amplia aplicación de la tecnología de inteligencia artificial en diversos campos ha acelerado el ritmo de exploración de la
8 información oculta detrás de grandes cantidades de datos. La gente espera utilizar métodos de minería de datos para realizar investigaciones
9 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
10 tecnología de minería de datos, la precisión de clasificación de alta precisión, los resultados de decisión intuitivos y la alta capacidad de
11 generalización lo convierten en un método más ideal de gestión de la educación superior. Con el objetivo de mejorar la sensibilidad del
12 procesamiento de datos y la clasificación del árbol de decisiones a los datos ruidosos, este documento propone mejoras correspondientes y
13 propone un estándar de selección de atributos de conjunto aproximado de precisión variable basado en la función de escala, que considera
14 tanto la precisión de aproximación ponderada como el valor del atributo. El número mejora la capacidad antiinterferente de los datos de ruido,
15 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
16 supresión, el soporte y la confianza en el proceso de poda previa del árbol, lo que simplifica la estructura del árbol. Los experimentos
17 comparativos en conjuntos de datos estándar muestran que el algoritmo mejorado propuesto en este documento es mejor que otros algoritmos
18 de árboles de decisión y puede realizar eficazmente la clasificación diferenciada de la gestión de la educación superior.

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

20 1. Introducción

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

32 La estrategia de diferenciación es una de las estrategias
33 básicas de competencia de las empresas propuestas por el
34 maestro de la estrategia Michael Porter. Esta teoría se aplica a
35 la investigación de las organizaciones e instituciones
36 educativas, y puede utilizarse para analizar la estrategia de
37 desarrollo de las universidades [2-4]. Con la creciente
38 diversificación de la educación superior, la competencia entre
39 las universidades de nuestro país se ha vuelto cada vez más
40 obvia, y se ha creado una situación en la que los fuertes se
41 hacen más fuertes y los débiles se debilitan. Esto también
42 demuestra que el modelo unilateral de gestión escolar de mil
43 escuelas no ha podido satisfacer las necesidades de la
44 economía de mercado, ni puede satisfacer el rápido desarrollo
45 de la educación superior [5-7]. Las universidades deben
46 reflexionar sobre su propia fuerza, posicionamiento y objetivos,
47 buscar un espacio de desarrollo adecuado en función de sus
48

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

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

56 La diversificación de las universidades es el consenso
 57 formado por los círculos de la educación superior en mi país
 58 que entran en la etapa de popularización de la educación
 59 superior. Sin embargo, la situación real de la educación
 60 superior de mi país está lejos de las exigencias de diversidad y
 61 características. En la actualidad, hay dos estrategias que
 62 prevalecen entre las universidades: una es expandir
 63 rápidamente la escala, fusionarse, ampliar la matrícula, adquirir
 64 terrenos y construir casas, y utilizar todos los medios para
 65 crecer más; la otra es fortalecer la fuerza de la investigación
 66 científica e introducir doctores, profesores e incluso
 67 académicos con una alta inversión [10]. Recompensar los logros
 68 de la investigación científica, lo que se manifiesta de manera
 69 destacada en el aumento del nivel de funcionamiento de las
 70 escuelas. Esto ha provocado una grave convergencia en el
 71 desarrollo de diferentes universidades, lo que resulta en
 72 similitudes y funciones de servicio superpuestas. Esto ha
 73 llevado a la escalada continua de la competencia viciosa entre
 74 las escuelas, y se desperdician, abusan y dejan sin usar valiosos
 75 recursos educativos. La distancia entre la educación y la
 76 realidad social está aumentando [11]. En China, la construcción
 77 de infraestructura universitaria, equipos de talentos de
 78 información y recursos de información educativa ha logrado un
 79 gran progreso, pero en realidad se ha centrado en cambiar la
 80 estructura de enseñanza tradicional centrada en el profesor y
 81 construir un nuevo tipo de enseñanza combinada de
 82 asignaturas principales. Menos del 5% de las escuelas están
 83 integradas con el objetivo de la estructura. Este era un
 84 problema común en el proceso de informatización de la
 85 educación en ese momento, y también era un cuello de botella
 86 que restringía el desarrollo en profundidad de la
 87 informatización de la educación en mi país [12]. En este
 88 contexto, el diseño del sistema de indicadores se ajusta justamente
 89 con la integración de la tecnología de la información y la
 90 enseñanza, discutiendo los puntos calientes y la urgencia de la
 91 demanda, y cambiando el enfoque de la investigación al
 92 proceso de aplicación de la información de la enseñanza.
 93

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

Por lo tanto, la universidad debe crear su propia competitividad
 101 única creando diferencias sobre la base de un estudio
 102 comparativo completo de los competidores y las necesidades
 103 sociales. Es decir, debemos hacer un uso completo de la
 104 estrategia de diferenciación para garantizar que la escuela
 105 tenga una capacidad sobresaliente para administrar una
 106 escuela y una excelente calidad de educación para tener la
 107 capacidad de competir continuamente. La gestión de la
 108 educación superior se muestra en la Figura 1.
 109

2. Clasificación diferenciada de la gestión de la 110 educación superior basada en inteligencia 111 artificial de precisión variable

112 La disciplina es el elemento básico de las universidades,
 113 el centro de reunión de talentos y la base para que
 114 profesores y estudiantes desempeñen sus funciones [13].
 115 Para que una escuela sea disuntiva, centrada, de marca,
 116 competitiva y competitiva es inseparable de disciplinas y
 117 grupos de disciplinas únicos y ventajosos. El académico
 118 estadounidense Fibre [14] señaló: "No importa dónde, el
 119 trabajo de la educación superior se compone de dos
 120 modos básicos: el cruce vertical y horizontal según las
 121 disciplinas e instituciones". En cierto sentido, la
 122 competitividad central de las universidades se concentra en
 123 sus disciplinas [15]. A nivel de construcción, las
 124 características de funcionamiento de las escuelas
 125 superiores y universidades se reflejan en última instancia
 126 en las características disciplinarias de la escuela. El
 127 indicador principal para juzgar si una universidad tiene
 128 competitividad central es su disciplina, y hay dos criterios
 129 para juzgar sus ventajas: uno es si está a la vanguardia del
 130 desarrollo científico y tecnológico del mundo; el otro es si
 131 se adapta a la tendencia del desarrollo social y económico
 132 [16]. Una universidad tiene su competitividad central solo
 133 cuando tiene una ventaja absoluta o relativa en una
 134 disciplina o grupo de disciplinas. La construcción
 135 disciplinaria, por un lado, debe esforzarse por fortalecer la
 136 construcción de disciplinas clave, utilizar disciplinas
 137 superiores para demostrar la radiación y promover el
 138 desarrollo coordinado de disciplinas relacionadas; por otro
 139 lado, también presta atención al desarrollo condensado de
 140 disciplinas y al cultivo de líderes disciplinarios. Existe un
 141 cierto grado de competencia entre los diversos
 142 departamentos dentro de la universidad debido al uso de
 143 los recursos limitados de la escuela. Es imposible que la
 144 escuela distribuya fondos y personal destacado de manera
 145 uniforme en toda la escuela [17, 18]. Por lo tanto, cuando
 146 las escuelas consideran la diferenciación, deben combinar
 147 la historia de la escuela.
 148

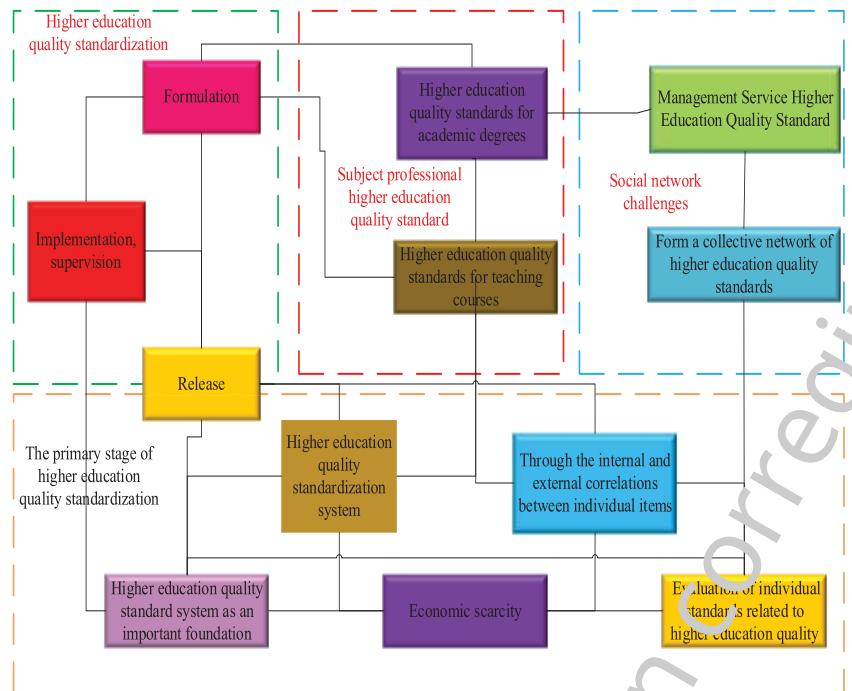


Fig. 1. Gestión de la educación superior.

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

160 2.1. Algoritmo de clasificación de árboles de decisión de 161 precisión variable basado en una función de escala

163 El árbol de decisión es un método simple pero ampliamente
 164 utilizado en modelos de clasificación y predicción. El árbol de
 165 decisión divide un gran conjunto de datos en partes más pequeñas
 166 [19–21]. Esta parte más pequeña del conjunto de datos se denomina
 167 nodo terminal. A cada nodo terminal se le asigna una etiqueta de
 168 clase como marca. Los nodos no terminales están compuestos por
 169 nodos raíz y otros nodos internos, e incluyen condiciones de prueba
 170 de atributos para distinguir registros de datos con diferentes
 171 características. Cuando el subconjunto de datos que se ha
 172 particionado no se puede partitionar aún más de acuerdo con los
 173 criterios de partición definidos previamente, se termina todo el
 174 proceso de partición.

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

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

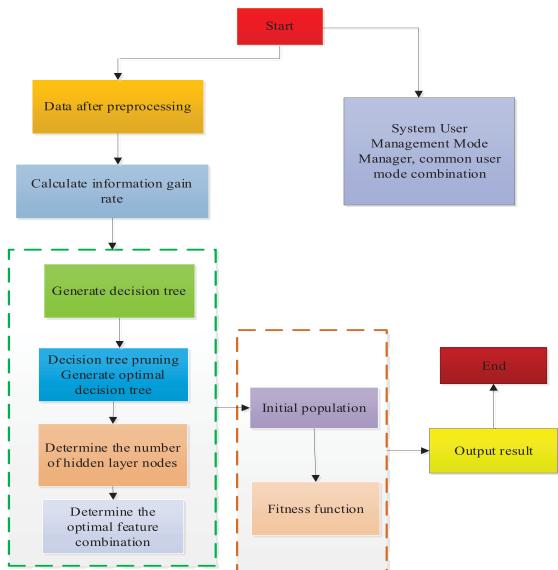


Fig. 2. Algoritmo de clasificación de árboles de decisión.

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

2.2 Teoría de precisión variable

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

224 Supongamos que la variable de escala $s=1, 2, \dots, N$, la
225 función de escala $f(s) \in [0.5, 1]$, la precisión de la división α
226 = 0.10, entonces $f(1) = 0.5, f(2) = 0.6, f(3) = 0.7, f(4) = 0.8, f(5) =$
227 0.9. Cuanto más precisa sea la división de la escala, más
228 detallada será la expresión del conocimiento.

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

$$B^*(f(s)) = \bigcup_{\{m_i | do(S, M_i) \leq 1 - f(s)\}} \quad (1)$$

$$B^*(f(s)) = \bigcup_{\{m_i | do(S, M_i) \geq f(s)\}} \quad (2)$$

Entre ellos, $do(Y/X) = 1 - |m_i \cap X|$ representa la tasa de clasificación errónea de clasificar Y como conjunto X .

2.3. Algoritmo mejorado de árboles 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 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 pade 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

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

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

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

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

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

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

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

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

$$exactitud = \frac{y_0}{y_0 + y_1} \quad (4)$$

319 Dónde $|condición_{muestra}^{regla}|$ representa el número
 320 Número de muestras de prueba y condiciones cumplidas en regla, $|$
 321 $condición_{regla}|$ representa el número total de condiciones en la
 322 regla, y_0 es el número de elementos correctos en el conjunto de
 323 datos de predicción, y_1 es el número de elementos incorrectos en el conjunto de
 324 datos de predicción, $y_0 + y_1$ es el número de etiquetas de clase correctas [20].

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

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

358 **3. Análisis de experimentos y resultados**359 **3.1 Análisis de casos**

Los ajustes de parámetros en el algoritmo MSVPRS DT son los siguientes: $f(s) = 0.8$, $\alpha = 0.5$, umbral del factor de supresión $del/min = 0.8$. Utilice la fórmula para calcular la importancia de cada atributo para obtener:

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

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

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

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

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 $t_{\text{último}} = \{1, 2, 3, 4, 8, 12, 14\}$,
 $t_{\text{último}} = \{5, 6, 7, 9, 10, 11, 13\}$, La clase Las etiquetas de estos dos subconjuntos son inconsistentes, para el subconjunto $t_{\text{último}}$, la confianza del subconjunto es $conferencia(normal) = GRAMO/7 > 0.8$, 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. $t_{\text{último}}$ Por lo tanto, se obtiene una rama con una longitud de arco normal y un nodo de hoja marcado como sí. Para otro subconjunto $t_{\text{último}}$, 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, Sig (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 $conf(construcción del equipo de talentos de informatización) = 1 > 0.8$, $conf(construcción de infraestructura) = 1 > 0.3$, el nivel de confianza del valor de atributo construcción de recursos de información $conf(construcción de recursos de información) = 0.5 < 0.8$, por lo que se obtiene una longitud de arco para la construcción de infraestructura y

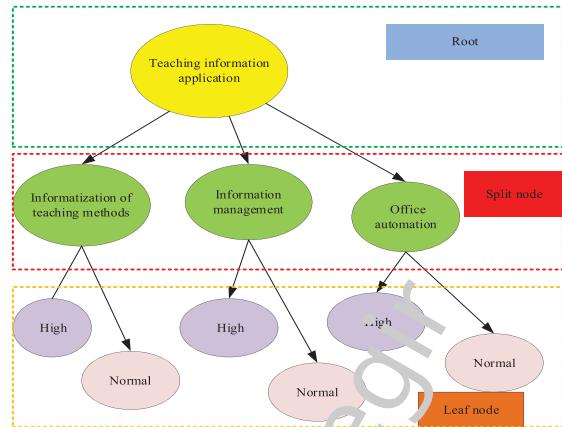


Fig. 3. Árbol de decisión generado por el algoritmo MSVPRS DT.

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 $f(s) = 0.9$, $f(s) = 1$ 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 $f(s) = 0$, 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%. $f(s) = 0.9$, $f(s) = 1$ tienen la misma estructura de árbol. La función de escala $f(s) = 0.9$, en comparación con $f(s) = 0.7$, es mejor que $f(s) = 0.7$ y $f(s) = 0.6$ en términos de precisión y estructura de árbol. Por lo tanto, para este ejemplo, la mejor función de escala es $f(s) = 0.9$.

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

Hay cuatro parámetros variables en el algoritmo MSVPRS DT: función de escala $f(s)$, umbral de confianza $minconf$, coeficiente de ajuste α , factor de supresión 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. $f(s)$ 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 MSVPRS DT 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

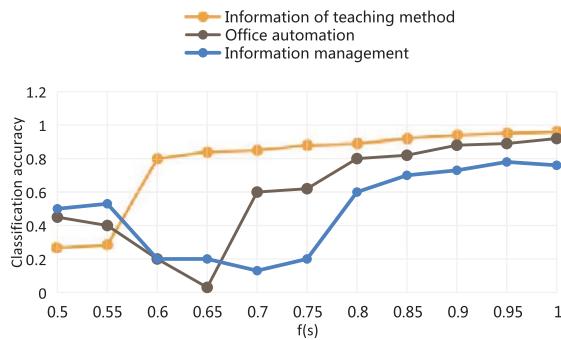


Fig. 4. Comparación de la precisión de la clasificación bajo diferentes funciones de escala $f(s)$.

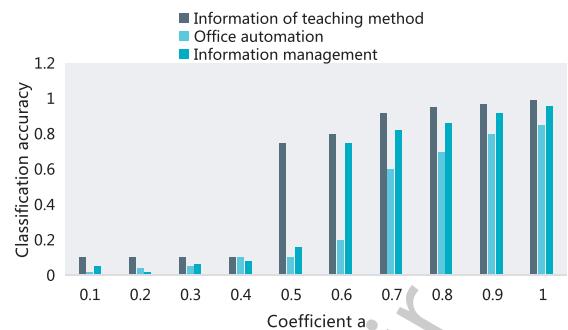


Fig. 6. Comparación de la precisión de la clasificación bajo diferentes coeficientes.

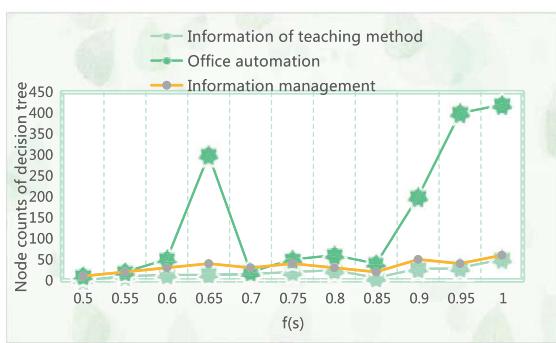


Fig. 5. Comparación del tamaño del árbol bajo diferentes funciones de escala $f(s)$.

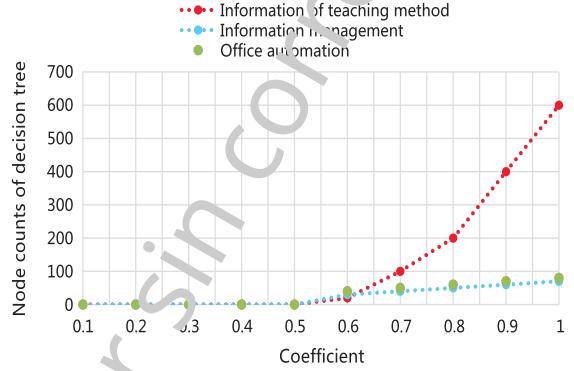


Fig. 7. Comparación del tamaño de los árboles bajo diferentes coeficientes.

426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
El árbol de expansión es proporcional a la función de escala. Cuanto mayor sea $f(s)$, tanto 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.

432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
Las figuras 6 y 7 muestran respectivamente la precisión de clasificación de diferentes coeficientes α en el conjunto de datos y el tamaño del árbol. La figura 6 muestra que cuando el coeficiente $\alpha < 0.5$, el cambio del coeficiente α tiene poco efecto en la precisión de la clasificación y la escala del árbol de expansión de los tres conjuntos de datos. $\alpha > 0.5$. A medida que aumenta el coeficiente α , la precisión de la clasificación y la escala del árbol de expansión del algoritmo MSVRSDT en el conjunto de datos tienen una tendencia a volverse más pequeñas. Cuando el coeficiente α aumenta, la precisión de la clasificación y la escala del árbol de expansión del algoritmo MSVRSDT en el conjunto de datos tienen una tendencia a volverse más pequeñas. $\alpha = 0.5$. Tanto la precisión de la clasificación como la escala del árbol de expansión en los tres conjuntos de datos.

444
445
446
447
448
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.

449
450
451
452
453
454
455
456
457
458
mediante la aproximación de la función objetivo discreta. Describir 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.

459
460
461
462
463
464
465
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.

466
467
468
469
470
471
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. $delmin > 0.9$. La tasa de precisión de la clasificación comienza a disminuir y la escala del árbol de expansión también comienza a disminuir.

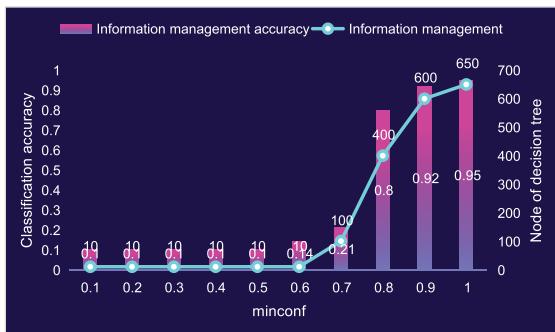


Fig. 8. Influencia del umbral de confianza en el algoritmo.



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.

483
484
485
486
487
488
489
490
491
492
493

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 la clasificación aproximada ponderable como el número de valores de atributo; en el método de poda de árboles de precisión, 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

494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531

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

535 **Expresiones de gratitud**

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

539 **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[], *Medicina natural* **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[], *Revista de informática en la educación superior* **32**(1) (2020), 57-73.
- [5] F. Farivar, MS Haghghi, 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 Steinhagen, 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[], *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[], *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* **81103722**.
- [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[], *Computación en clúster* **22**(3) (2019), 7435-7445.
- [11] G. Manogaran, S. Baskar, PM Shakeel, N. Chikamkurti 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[], *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[], *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[], *Computación en clúster* **22**(6) (2019), 14559-14581.
- [15] U. Fiore, A. De Santis, F. Perla, et al., Uso de redes generativas antagónicas para mejorar la eficacia de la clasificación en la detección de fraudes con tarjetas de crédito[], *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[], *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[], *Acceso IEEE* **14** (2016), 2379-2387.
- [18] YJ Yang y CS Bang, Aplicación de inteligencia artificial en gastroenterología[], *Revista Mundial de Gastroenterología* **25**(14) (2019), 16-19.
- [19] H. Xiao, B. Muthu y SN Kandasamy, Corrección de: Inteligencia artificial con robótica para la industria manufacturera avanzada utilizando un modelo de programación de números enteros mixtos cristalizado 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-Conf. Iberamericana internacional sobre comunicaciones del IEEE 2020* (CIC), Dublín, Irlanda, 2020, págs. 1-6, doi: 10.1109/IC-40277.2020.9148850.
- [21] C. Kittranongwong, AS Bomba, U. Baber, et al., Dirección futura para el uso de inteligencia artificial para predecir y controlar la hipertensión[], *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[], *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[], *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[], *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[], *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[], *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[], *Revista británica de radiología* **92**(3) (2019), 38-39.

- 652 [28] S. Zubaidah, NM Fuad, S. Mahanal, et al., Mejora de las habilidades de 661
653 pensamiento creativo de los estudiantes a través de la investigación 662
654 científica diferenciada integrada con mapas mentales[], *Revista de 663
655 educación científica turca*14(4) (2017), 77-91.
656 [29] MF Byrne, N. Chapados, F. Soudan, et al., Diferenciación en tiempo real de 664
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. Sàncchez-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.

Prueba de autor sin corregir