

ایجاد قواعد تصمیم با استفاده از الگوریتم ژنتیک بر اساس خوشه‌بندی

مهدی نصیری^۱، علی هادیان^۲، بهروز مینایی^۳

چکیده

تصمیم‌گیری یکی از کارهای مهمی است که می‌توان روی پایگاه داده‌های بزرگ انجام داد. در این روش درخت تصمیم‌گیری با استفاده از الگوریتم ژنتیک ایجاد می‌شود. این تصمیم‌گیری بر اساس داده‌های خوشه‌بندی شده انجام می‌گیرد. طول کروموزوم‌ها در این روش $3n$ است و هر ژن ۲ بخش دارد. بخش اول عددی بین ۱ تا تعداد فیلدها است که بیانگر شماره فیلد است و بخش دوم یک رشته ۲ بیتی باینری است. کروموزوم‌هایی بصورت تصادفی ایجاد می‌شوند. بعد از ترکیب و جهش بر روی آنها یکسری از آنها برای تشکیل درخت بر اساس تابع ارزش انتخاب می‌شوند. بهترین کروموزوم رده‌بند اصلی در نظر گرفته می‌شود و بقیه کروموزوم‌ها با توجه به حالت بهتر نسبت به ریشه درخت درخت تصمیم‌گیری قرار می‌گیرند. هر کروموزوم یک مسیر درخت از ریشه تا برگ است و کروموزوم‌ها به ترتیب ارزش آنها در تشکیل درخت تصمیم‌گیر شرکت می‌کنند. نتایج نشان از افزایش سرعت استخراج قواعد تصمیم بر اساس داده‌های خوشه‌بندی شده است.

کلمات کلیدی

قواعد تصمیم‌گیری، الگوریتم ژنتیک، رده‌بندی، تابع ارزش، خوشه‌بندی

۱. مقدمه

یک درخت تصمیم یک ساختار سلسله‌مراتبی می‌باشد که در آن، گره‌های میانی برای آزمون یک خصیصه^۱ به کار می‌روند. شاخه‌ها نشان‌گر خروجی آزمون بوده، برگ‌ها برچسب رده^۲ و یا توزیع بر چسب رده را مشخص می‌نمایند. نکات اساسی برای هر درخت تصمیم [Kennedy۹۷] عوامل ملاک استفاده شده برای ساخت درخت، (یعنی کدام متغیر باید برای شکستن انتخاب گردد و این متغیر چگونه باید شکسته شود؟) ملاک برای متوقف کردن رشد درخت (یعنی چه موقعی باید عمل شاخه شاخه شدن یک نود باید متوقف شود؟) و نحوه هرس شاخه‌های درخت بدست آمده برای بیشترین کارایی در رده‌بندی است. در این تحقیق یک روش تولید درخت تصمیم با استفاده از الگوریتم ژنتیک ارائه شده است. در بخش ۲ درخت تصمیم، بخش ۳ الگوریتم ژنتیک و در بخش ۴ روش پیشنهادی ارائه می‌شود. در بخش ۵ نیز نتایج و کارهای آینده و در بخش آخر منابع ارائه می‌گردد.

۲. درخت تصمیم

یک درخت تصمیم، عبارت است از یک ساختار درخت‌گونه به علاوه برخی رویه‌های استنتاج از روی آن. علاوه بر این، درخت‌های یادگرفته شده می‌توانند در قالب مجموعه‌ای از قوانین «اگر-آنگاه» بازنمایانده شوند تا قابلیت فهم^۳ آنها توسط انسان، ارتقاء پیدا نماید. یادگیری درخت تصمیم یکی از متداول‌ترین و عملی‌ترین روش‌های یادگیری استقرایی^۴ است. این رهیافت، روشی برای تقریب توابع مقدار-گسسته^۵ می‌باشد [۳]. دامنه ارزش خصیصه‌ها، در حالت کلی می‌تواند دو نوع متفاوت داشته باشد: گسسته و پیوسته. در بسیاری از الگوریتم‌ها لازم است که داده‌ها با مقادیر گسسته شرح داده شوند. پیش از اجرای چنین الگوریتم‌هایی، می‌بایست یک فرآیند گسسته‌سازی^۶ روی دامنه مقادیر خصیصه‌های مقدار-پیوسته صورت گیرد. این امر در حالت کلی نیازمند نوعی افراز^۷ یا پوشش^۸ می‌باشد. روش‌های مختلفی برای گسسته‌سازی پیشنهاد گردیده‌اند؛ به طور کلی، این روش‌ها از سه جنبه زیر قابل تقسیم‌بندی هستند [۴]:

^۱ دانشجوی کارشناسی ارشد هوش مصنوعی دانشگاه علم و صنعت ایران nasiri@comp.iust.ac.ir

^۲ دانشجوی کارشناسی مهندسی نرم‌افزار دانشگاه علم و صنعت ایران hadian@comp.iust.ac.ir

^۳ استادیار دانشکده مهندسی کامپیوتر دانشگاه علم و صنعت ایران b_minaei@iust.ac.ir

- محلی در مقابل سراسری: یک روش محلی، عمل گسسته‌سازی را در یک ناحیه محلی شده از فضای نمونه‌ای (یعنی زیرمجموعه‌ای از نمونه‌ها) انجام می‌دهد. در حالی که یک روش گسسته‌سازی سراسری، تمامی فضای نمونه‌ای را برای گسسته‌سازی به کار می‌گیرد [۵].
 - تحت نظارت^۹ در مقابل بدون نظارت^{۱۰}: بسیاری از روش‌های گسسته‌سازی اطلاعات رده‌ای را در حین فرآیند گسسته‌سازی مورد استفاده قرار نمی‌دهند. این نوع روش‌ها، روش‌های بدون نظارت نامیده می‌شوند. در مقابل، روش‌هایی که از چنین اطلاعاتی برای گسسته‌سازی استفاده می‌نمایند [۶]، به روش‌های تحت نظارت موسوم هستند.
 - ایستا در مقابل پویا: بسیاری از روش‌های گسسته‌سازی نیاز به یک پارامتر k دارند، که نمایانگر بیشینه تعداد بازه‌های افراز در یک خصیصه گسسته‌سازی شده می‌باشد. در مقابل، روش‌های پویا در فضای مقادیر ممکن k ، برای همه خصیصه‌ها به طور همزمان جستجو نموده، به این وسیله وابستگی‌های مابین خصیصه‌ها را دریافته، و در فرآیند گسسته‌سازی مورد بهره‌برداری قرار می‌دهند [۷].
- الگوریتم‌های یادگیری درخت تصمیم به طور کلی بر اساس روش «تقسیم‌بندی بازگشتی»^{۱۱} کار می‌کنند. ایده بنیادی این روش‌ها عبارتست از: افراز فضای نمونه‌ای در یک اسلوب داده‌گرا^{۱۲} و نمایش این افراز به صورت یک درخت [۱]. بیشتر الگوریتم‌هایی که برای یادگیری درخت‌های تصمیم توسعه داده شده‌اند، گونه‌هایی از یک الگوریتم پایه هستند که یک جستجوی حریصانه بالا به پایین را در فضای کلیه درخت‌های تصمیم هدایت می‌کنند. اساس این الگوریتم‌ها انتخاب‌های بهینه محلی است که در هر گره صورت می‌گیرد. این رهیافت توسط الگوریتم‌هایی از قبیل ID۳ [۸]، CART [۹] و C۴.۵ [۱۰] به خدمت گرفته شده است. ثابت شده است که ماهیت حریصانه این الگوریتم‌ها اغلب منجر به تولید درخت‌های غیربهینه می‌شود. اگرچه [۱۱] ثابت کرده است که مسأله ساخت درخت تصمیم بهینه یک مسأله NP-Complete می‌باشد، اما تکنیک‌های فراوانی برای حرکت به سمت درخت‌های بهینه‌تر ارائه شده است.
- در استقرای درخت تصمیم همواره مطلوب است که درخت ساخته شده حتی الامکان کوچک باشد. اندازه درخت‌های تصمیم به طور عمده بستگی به چگونگی انتخاب خصیصه‌ای دارد که در هر گره برای تقسیم نمونه‌های آن مورد استفاده قرار می‌گیرد. چنین خصیصه‌ای خصیصه انشعاب^{۱۳} نامیده می‌شود. برای اندازه‌گیری کیفیت خصیصه‌ها، به عنوان خصیصه انشعاب، چندین معیار توسعه داده شده‌اند. معیارهای متداول عبارتند از آنتروپی (بهره اطلاعاتی)، اندیس Gini [۱۲]، معیار χ^2 [۱۳، ۱۴] و نظایر آن. با وجود حجم زیاد مطالعات مقایسه‌ای [۱۵]، آزمایش‌های بسیار کمی وجود داشته‌اند که نتیجه‌گیری نمایند یکی از این معیارها بهتر از سایرین عمل می‌کند. اغلب معیاری که در مورد برخی از مسائل خیلی خوب عمل می‌کند، در زمینه حل انواع دیگری از مسائل کارایی پائین‌تری نسبت به معیارهای دیگر از خود نشان می‌دهد.
- اکثر الگوریتم‌های ساخت درخت تصمیم، هر شاخه از درخت را تا عمقی که برای رده‌بندی کامل نمونه‌های آموزشی لازم باشد، توسعه می‌دهد. اگر چه این استراتژی در مواقعی معقول به نظر می‌رسد، ولیکن زمانی که در داده‌ها اغتشاش وجود داشته باشد یا هنگامی که تعداد نمونه‌های آموزشی به اندازه‌ای کوچک است که نمی‌تواند به نحو مناسبی تابع هدف را بازنمایی کند، مشکلاتی بروز می‌نماید. در هر یک از این دو حالت، این الگوریتم ساده ممکن است درختانی تولید کند که روی نمونه‌های آموزشی فوق برازش^{۱۴} حاصل کرده باشند. راه‌حل‌های مختلفی برای مقابله با پدیده فوق برازش وجود دارند. این‌ها را می‌توان به دو دسته کلی زیر تقسیم کرد:
- روش‌هایی که رشد درخت را زودتر از موعد متوقف می‌کنند.
 - روش‌هایی که اجازه می‌دهند، فوق برازش حاصل شود، آنگاه درخت را هرس می‌کنند.
- روش‌های دسته نخست را هرس زمان ساخت^{۱۵}، و روش‌های دسته دوم را هرس بعد از ساخت^{۱۶} می‌نامند.
- مکانیزم استنتاج از روی درخت تصمیم به این صورت انجام می‌گیرد که در هنگام تصمیم‌گیری برای یک نمونه جدید، مقدار خصیصه انشعاب گره ریشه در آن نمونه، با مقادیر خصیصه در شاخه‌ها مطابقت داده می‌شود. این فرآیند برای زیردرختی که در زیر شاخه‌ای قرار دارد که شرط آن ارضا شده است، تکرار می‌شود و به همین ترتیب ادامه می‌یابد تا اینکه به یک گره برگ منتهی شود. برچسب این گره برگ به عنوان کلاس نمونه مفروض در نظر گرفته خواهد شد.

۳. الگوریتم ژنتیک

الگوریتم ژنتیک^{۱۷} رهیافتی است که تکامل طبیعی موجودات را الگو قرار می‌دهد. این روش تقلیدی از فرایند تکامل با استفاده از الگوریتم‌های کامپیوتری است. اساسی‌ترین اصل تکامل، وراثت است. هر نسل، خصوصیات نسل قبلی را به ارث می‌برد و به نسل بعد انتقال می‌دهد. این انتقال خصوصیات از نسلی به نسل بعد توسط ژن‌ها صورت می‌گیرد. جهانی که در آن زندگی می‌کنیم دائماً در حال تغییر است. برای بقا در این سیستم پویا، افراد باید توانایی داشته باشند که خود را با محیط، سازگار کنند. سازگاری^{۱۸} تعیین می‌کند که آن موجود چه مقدار زنده خواهد ماند و چقدر شانس

دارد تا ژن‌های خود را به نسل بعد انتقال دهد. در تکامل بیولوژیکی، فقط برنده‌ها هستند که می‌توانند در فرایند تکامل شرکت کنند. خصوصیات هر موجود زنده در ژن‌هایش، کدگذاری شده است و طی فرایند وراثت، این ژن‌ها به فرزندان^{۱۹} منتقل می‌شوند. مبتکر الگوریتم ژنتیک جان هلند [Holland ۱۹۷۵]^{۲۰} «در دهه هفتاد میلادی با الهام گرفتن از «ویژگی‌های تئوری تکامل، الگوریتم جستجویی ابداع کرد که در این الگوریتم از همان اصولی که طبیعت فرایند تکامل را روی نمادهای ژنی انجام می‌دهد، برای تکامل جواب‌های مربوط به حل‌های یک مساله بهینه‌سازی استفاده می‌کند. فرایند با یک جمعیت اولیه تصادفی از جواب‌های ممکن شروع می‌شود. هریک از جواب‌ها توسط یک ساختار رشته‌ای از بیت‌ها که مقدار کدگذاری شده متغیرهای تصمیم را در بردارند، نشان داده می‌شوند. سپس با تشکیل خانواده اولیه و ارزیابی هر یک از رشته‌ها، افراد مناسب برای تشکیل خانواده بعدی انتخاب می‌شوند. جواب‌های جدید از خانواده جواب‌های اولیه با تغییر دادن ساختار رشته‌ها توسط عملگرهای الگوریتم ژنتیک تولید می‌شوند. رشته‌های جدید توسط روند طراحی الهام گرفته از مکانیزم ژنتیک طبیعی تولید می‌شوند. سپس مقدار برازندگی رشته‌های جدید با توجه به تابع هدف مسأله مورد نظر ارزیابی می‌شود. این روند موجب بهبود مداوم برازندگی خانواده حل‌ها شده و تا زمانی که حل‌ها همگرا شوند، تکرار می‌شود. دو جنبه مهم در الگوریتم ژنتیک وجود دارند که دائما جواب‌ها را آشفته کرده و مجال خروج از بهینه‌های موضعی را فراهم می‌آورند. یکی از این جنبه‌ها آمیزش است که GA از آن برای تولید جواب استفاده می‌کند. جنبه دیگر که عملگر جهش نام دارد، قادر است مقادیر جدیدی به بیت‌ها بدهد که در گروه والدین وجود نداشته است. عملگر جهش کمک می‌کند که تنوع ژنتیک باقی بماند و جستجو به نواحی جدیدی برسد.

بطور کلی می‌توان روش را بدین صورت شرح داد که در نخستین مرحله الگوریتم ژنتیک، جمعیتی از کروموزوم‌ها به تعداد معین و بطور تصادفی تولید می‌شوند. هر کروموزوم بیانگر یک جواب از فضای جستجو است و فرد نام دارد. مجموعه این افراد، جمعیت یا نسل فعلی نام دارند. به هر شخص، برازندگی بر اساس مقدار تعیین شده توسط تابع هدف تعلق می‌گیرد. از این مقدار برای هدایت فرایند انتخاب به سمت اشخاص مناسب‌تر استفاده می‌گردد.

افراد با برازندگی بالا نسبت به کل جمعیت، احتمال بیشتری برای انتخاب شدن برای تولید مثل دارند و در مقابل، اشخاص با برازندگی کمتر احتمال انتخاب کمتری دارند. پس از تعیین مقدار برازندگی اعضای جمعیت، می‌توان آنها را با احتمالی متناظر با برازندگی نسبی‌شان انتخاب کرد و برای تولید نسل بعد ترکیب نمود. در مرحله بعد نوبت به اعمال عملگرهای آمیزش و جهش می‌رسد. سپس در صورت لزوم رشته‌های اشخاص جمعیت، کدگشایی شده و تابع هدف ارزیابی می‌شود و مقدار برازندگی هر فرد تعیین می‌گردد و اشخاص مطابق با مقدار برازندگی برای جفت‌گیری انتخاب می‌شوند و فرایند به همین ترتیب تا تولید نسل بعد ادامه می‌یابد. الگوریتم ژنتیک هنگامی که برخی ضوابط مانند تعداد معینی تولید نسل و یا میانگین انحراف معیار عملکرد اشخاص جمعیت تأمین شود، به پایان می‌رسد.

۴. روش پیشنهادی

چون محاسبه بخصوص زمانی که پایگاه داده برای پیمایش نمی‌تواند کاملاً در حافظه اصلی تغذیه شود هزینه زیادی دارد از الگوریتم ژنتیک برای کاهش هزینه استفاده شده است. این روش روی داده‌های دودویی قابل اجرا است. در این روش یکی از کارهای مفیدی که می‌توان برای داده‌کاوی پیشنهاد داد افزایش سرعت است. از آنجا که داده‌ها معمولاً برای استخراج زیاد است، کاهش پیمایش پایگاه داده می‌تواند در این امر موثر باشد. چون هر بار که حلقه برای تولید قواعد اجرا می‌شود، باید یک‌بار کل پایگاه داده باید پیمایش شود.

در این روش مانند برخی از روش‌ها فقط یک‌بار پایگاه داده را پیمایش می‌کند. این روش ابتدا جدول‌هایی از مجموعه‌های خوشه‌بندی شده را ایجاد می‌کند. K خوشه ایجاد می‌شود که K طول حداکثر تراکنش است. روش بدین صورت است که داده‌ها به تعداد حداکثر طول داده‌ها خوشه‌بندی می‌شوند. مثلاً اگر حداکثر طول داده‌ها ۲۰ باشد، داده‌ها به ۲۰ خوشه تقسیم می‌شود. در خوشه اول داده‌هایی که یک اقلام حداکثر دارند، در خوشه دوم داده‌های با طول ۲ قلم حداکثر و در خوشه ۲۰ داده‌های با صفت حداکثر ۲۰ تا قرار می‌گیرند. این کار باعث می‌شود فقط یک‌بار پایگاه داده پیمایش شود. همچنین مثلاً اگر یک قاعده دارای ۵ صفت باشد، برای محاسبه تابع ارزش فقط جدول‌های بیشتر از ۵ قلم را جستجو می‌کند. روش پیشنهادی باعث پیمایش یک‌بار داده‌ها و در نتیجه افزایش سرعت می‌شود. درخت تصمیم‌گیری با استفاده از الگوریتم ژنتیک ایجاد می‌شود. طول کروموزوم‌ها در این روش ۳ برابر تعداد فیلدهای تصمیم‌گیری است و هر ژن ۲ بخش دارد. بخش اول عددی بین ۱ تا تعداد فیلدها است که بیانگر شماره فیلد است و بخش دوم یک رشته ۲ بیتی باینری است. بعنوان مثال قرار است که برای یک مجموعه داده با ۶ فیلد که یکی از آنها نوع رده را مشخص می‌کند و ۵ فیلد دیگر، فیلدهای تصمیم هستند. در این مثال طول هر کروموزوم ۱۵ است که بیت‌های شماره ۱۳، ۱۴، ۱۵، ۱۶، ۱۷، ۱۸، ۱۹، ۲۰، ۲۱، ۲۲، ۲۳، ۲۴، ۲۵، ۲۶، ۲۷، ۲۸، ۲۹، ۳۰، ۳۱، ۳۲، ۳۳، ۳۴، ۳۵، ۳۶، ۳۷، ۳۸، ۳۹، ۴۰، ۴۱، ۴۲، ۴۳، ۴۴، ۴۵، ۴۶، ۴۷، ۴۸، ۴۹، ۵۰، ۵۱، ۵۲، ۵۳، ۵۴، ۵۵، ۵۶، ۵۷، ۵۸، ۵۹، ۶۰، ۶۱، ۶۲، ۶۳، ۶۴، ۶۵، ۶۶، ۶۷، ۶۸، ۶۹، ۷۰، ۷۱، ۷۲، ۷۳، ۷۴، ۷۵، ۷۶، ۷۷، ۷۸، ۷۹، ۸۰، ۸۱، ۸۲، ۸۳، ۸۴، ۸۵، ۸۶، ۸۷، ۸۸، ۸۹، ۹۰، ۹۱، ۹۲، ۹۳، ۹۴، ۹۵، ۹۶، ۹۷، ۹۸، ۹۹، ۱۰۰، ۱۰۱، ۱۰۲، ۱۰۳، ۱۰۴، ۱۰۵، ۱۰۶، ۱۰۷، ۱۰۸، ۱۰۹، ۱۱۰، ۱۱۱، ۱۱۲، ۱۱۳، ۱۱۴، ۱۱۵، ۱۱۶، ۱۱۷، ۱۱۸، ۱۱۹، ۱۲۰، ۱۲۱، ۱۲۲، ۱۲۳، ۱۲۴، ۱۲۵، ۱۲۶، ۱۲۷، ۱۲۸، ۱۲۹، ۱۳۰، ۱۳۱، ۱۳۲، ۱۳۳، ۱۳۴، ۱۳۵، ۱۳۶، ۱۳۷، ۱۳۸، ۱۳۹، ۱۴۰، ۱۴۱، ۱۴۲، ۱۴۳، ۱۴۴، ۱۴۵، ۱۴۶، ۱۴۷، ۱۴۸، ۱۴۹، ۱۵۰، ۱۵۱، ۱۵۲، ۱۵۳، ۱۵۴، ۱۵۵، ۱۵۶، ۱۵۷، ۱۵۸، ۱۵۹، ۱۶۰، ۱۶۱، ۱۶۲، ۱۶۳، ۱۶۴، ۱۶۵، ۱۶۶، ۱۶۷، ۱۶۸، ۱۶۹، ۱۷۰، ۱۷۱، ۱۷۲، ۱۷۳، ۱۷۴، ۱۷۵، ۱۷۶، ۱۷۷، ۱۷۸، ۱۷۹، ۱۸۰، ۱۸۱، ۱۸۲، ۱۸۳، ۱۸۴، ۱۸۵، ۱۸۶، ۱۸۷، ۱۸۸، ۱۸۹، ۱۹۰، ۱۹۱، ۱۹۲، ۱۹۳، ۱۹۴، ۱۹۵، ۱۹۶، ۱۹۷، ۱۹۸، ۱۹۹، ۲۰۰، ۲۰۱، ۲۰۲، ۲۰۳، ۲۰۴، ۲۰۵، ۲۰۶، ۲۰۷، ۲۰۸، ۲۰۹، ۲۱۰، ۲۱۱، ۲۱۲، ۲۱۳، ۲۱۴، ۲۱۵، ۲۱۶، ۲۱۷، ۲۱۸، ۲۱۹، ۲۲۰، ۲۲۱، ۲۲۲، ۲۲۳، ۲۲۴، ۲۲۵، ۲۲۶، ۲۲۷، ۲۲۸، ۲۲۹، ۲۳۰، ۲۳۱، ۲۳۲، ۲۳۳، ۲۳۴، ۲۳۵، ۲۳۶، ۲۳۷، ۲۳۸، ۲۳۹، ۲۴۰، ۲۴۱، ۲۴۲، ۲۴۳، ۲۴۴، ۲۴۵، ۲۴۶، ۲۴۷، ۲۴۸، ۲۴۹، ۲۵۰، ۲۵۱، ۲۵۲، ۲۵۳، ۲۵۴، ۲۵۵، ۲۵۶، ۲۵۷، ۲۵۸، ۲۵۹، ۲۶۰، ۲۶۱، ۲۶۲، ۲۶۳، ۲۶۴، ۲۶۵، ۲۶۶، ۲۶۷، ۲۶۸، ۲۶۹، ۲۷۰، ۲۷۱، ۲۷۲، ۲۷۳، ۲۷۴، ۲۷۵، ۲۷۶، ۲۷۷، ۲۷۸، ۲۷۹، ۲۸۰، ۲۸۱، ۲۸۲، ۲۸۳، ۲۸۴، ۲۸۵، ۲۸۶، ۲۸۷، ۲۸۸، ۲۸۹، ۲۹۰، ۲۹۱، ۲۹۲، ۲۹۳، ۲۹۴، ۲۹۵، ۲۹۶، ۲۹۷، ۲۹۸، ۲۹۹، ۳۰۰، ۳۰۱، ۳۰۲، ۳۰۳، ۳۰۴، ۳۰۵، ۳۰۶، ۳۰۷، ۳۰۸، ۳۰۹، ۳۱۰، ۳۱۱، ۳۱۲، ۳۱۳، ۳۱۴، ۳۱۵، ۳۱۶، ۳۱۷، ۳۱۸، ۳۱۹، ۳۲۰، ۳۲۱، ۳۲۲، ۳۲۳، ۳۲۴، ۳۲۵، ۳۲۶، ۳۲۷، ۳۲۸، ۳۲۹، ۳۳۰، ۳۳۱، ۳۳۲، ۳۳۳، ۳۳۴، ۳۳۵، ۳۳۶، ۳۳۷، ۳۳۸، ۳۳۹، ۳۴۰، ۳۴۱، ۳۴۲، ۳۴۳، ۳۴۴، ۳۴۵، ۳۴۶، ۳۴۷، ۳۴۸، ۳۴۹، ۳۵۰، ۳۵۱، ۳۵۲، ۳۵۳، ۳۵۴، ۳۵۵، ۳۵۶، ۳۵۷، ۳۵۸، ۳۵۹، ۳۶۰، ۳۶۱، ۳۶۲، ۳۶۳، ۳۶۴، ۳۶۵، ۳۶۶، ۳۶۷، ۳۶۸، ۳۶۹، ۳۷۰، ۳۷۱، ۳۷۲، ۳۷۳، ۳۷۴، ۳۷۵، ۳۷۶، ۳۷۷، ۳۷۸، ۳۷۹، ۳۸۰، ۳۸۱، ۳۸۲، ۳۸۳، ۳۸۴، ۳۸۵، ۳۸۶، ۳۸۷، ۳۸۸، ۳۸۹، ۳۹۰، ۳۹۱، ۳۹۲، ۳۹۳، ۳۹۴، ۳۹۵، ۳۹۶، ۳۹۷، ۳۹۸، ۳۹۹، ۴۰۰، ۴۰۱، ۴۰۲، ۴۰۳، ۴۰۴، ۴۰۵، ۴۰۶، ۴۰۷، ۴۰۸، ۴۰۹، ۴۱۰، ۴۱۱، ۴۱۲، ۴۱۳، ۴۱۴، ۴۱۵، ۴۱۶، ۴۱۷، ۴۱۸، ۴۱۹، ۴۲۰، ۴۲۱، ۴۲۲، ۴۲۳، ۴۲۴، ۴۲۵، ۴۲۶، ۴۲۷، ۴۲۸، ۴۲۹، ۴۳۰، ۴۳۱، ۴۳۲، ۴۳۳، ۴۳۴، ۴۳۵، ۴۳۶، ۴۳۷، ۴۳۸، ۴۳۹، ۴۴۰، ۴۴۱، ۴۴۲، ۴۴۳، ۴۴۴، ۴۴۵، ۴۴۶، ۴۴۷، ۴۴۸، ۴۴۹، ۴۵۰، ۴۵۱، ۴۵۲، ۴۵۳، ۴۵۴، ۴۵۵، ۴۵۶، ۴۵۷، ۴۵۸، ۴۵۹، ۴۶۰، ۴۶۱، ۴۶۲، ۴۶۳، ۴۶۴، ۴۶۵، ۴۶۶، ۴۶۷، ۴۶۸، ۴۶۹، ۴۷۰، ۴۷۱، ۴۷۲، ۴۷۳، ۴۷۴، ۴۷۵، ۴۷۶، ۴۷۷، ۴۷۸، ۴۷۹، ۴۸۰، ۴۸۱، ۴۸۲، ۴۸۳، ۴۸۴، ۴۸۵، ۴۸۶، ۴۸۷، ۴۸۸، ۴۸۹، ۴۹۰، ۴۹۱، ۴۹۲، ۴۹۳، ۴۹۴، ۴۹۵، ۴۹۶، ۴۹۷، ۴۹۸، ۴۹۹، ۵۰۰، ۵۰۱، ۵۰۲، ۵۰۳، ۵۰۴، ۵۰۵، ۵۰۶، ۵۰۷، ۵۰۸، ۵۰۹، ۵۱۰، ۵۱۱، ۵۱۲، ۵۱۳، ۵۱۴، ۵۱۵، ۵۱۶، ۵۱۷، ۵۱۸، ۵۱۹، ۵۲۰، ۵۲۱، ۵۲۲، ۵۲۳، ۵۲۴، ۵۲۵، ۵۲۶، ۵۲۷، ۵۲۸، ۵۲۹، ۵۳۰، ۵۳۱، ۵۳۲، ۵۳۳، ۵۳۴، ۵۳۵، ۵۳۶، ۵۳۷، ۵۳۸، ۵۳۹، ۵۴۰، ۵۴۱، ۵۴۲، ۵۴۳، ۵۴۴، ۵۴۵، ۵۴۶، ۵۴۷، ۵۴۸، ۵۴۹، ۵۵۰، ۵۵۱، ۵۵۲، ۵۵۳، ۵۵۴، ۵۵۵، ۵۵۶، ۵۵۷، ۵۵۸، ۵۵۹، ۵۶۰، ۵۶۱، ۵۶۲، ۵۶۳، ۵۶۴، ۵۶۵، ۵۶۶، ۵۶۷، ۵۶۸، ۵۶۹، ۵۷۰، ۵۷۱، ۵۷۲، ۵۷۳، ۵۷۴، ۵۷۵، ۵۷۶، ۵۷۷، ۵۷۸، ۵۷۹، ۵۸۰، ۵۸۱، ۵۸۲، ۵۸۳، ۵۸۴، ۵۸۵، ۵۸۶، ۵۸۷، ۵۸۸، ۵۸۹، ۵۹۰، ۵۹۱، ۵۹۲، ۵۹۳، ۵۹۴، ۵۹۵، ۵۹۶، ۵۹۷، ۵۹۸، ۵۹۹، ۶۰۰، ۶۰۱، ۶۰۲، ۶۰۳، ۶۰۴، ۶۰۵، ۶۰۶، ۶۰۷، ۶۰۸، ۶۰۹، ۶۱۰، ۶۱۱، ۶۱۲، ۶۱۳، ۶۱۴، ۶۱۵، ۶۱۶، ۶۱۷، ۶۱۸، ۶۱۹، ۶۲۰، ۶۲۱، ۶۲۲، ۶۲۳، ۶۲۴، ۶۲۵، ۶۲۶، ۶۲۷، ۶۲۸، ۶۲۹، ۶۳۰، ۶۳۱، ۶۳۲، ۶۳۳، ۶۳۴، ۶۳۵، ۶۳۶، ۶۳۷، ۶۳۸، ۶۳۹، ۶۴۰، ۶۴۱، ۶۴۲، ۶۴۳، ۶۴۴، ۶۴۵، ۶۴۶، ۶۴۷، ۶۴۸، ۶۴۹، ۶۵۰، ۶۵۱، ۶۵۲، ۶۵۳، ۶۵۴، ۶۵۵، ۶۵۶، ۶۵۷، ۶۵۸، ۶۵۹، ۶۶۰، ۶۶۱، ۶۶۲، ۶۶۳، ۶۶۴، ۶۶۵، ۶۶۶، ۶۶۷، ۶۶۸، ۶۶۹، ۶۷۰، ۶۷۱، ۶۷۲، ۶۷۳، ۶۷۴، ۶۷۵، ۶۷۶، ۶۷۷، ۶۷۸، ۶۷۹، ۶۸۰، ۶۸۱، ۶۸۲، ۶۸۳، ۶۸۴، ۶۸۵، ۶۸۶، ۶۸۷، ۶۸۸، ۶۸۹، ۶۹۰، ۶۹۱، ۶۹۲، ۶۹۳، ۶۹۴، ۶۹۵، ۶۹۶، ۶۹۷، ۶۹۸، ۶۹۹، ۷۰۰، ۷۰۱، ۷۰۲، ۷۰۳، ۷۰۴، ۷۰۵، ۷۰۶، ۷۰۷، ۷۰۸، ۷۰۹، ۷۱۰، ۷۱۱، ۷۱۲، ۷۱۳، ۷۱۴، ۷۱۵، ۷۱۶، ۷۱۷، ۷۱۸، ۷۱۹، ۷۲۰، ۷۲۱، ۷۲۲، ۷۲۳، ۷۲۴، ۷۲۵، ۷۲۶، ۷۲۷، ۷۲۸، ۷۲۹، ۷۳۰، ۷۳۱، ۷۳۲، ۷۳۳، ۷۳۴، ۷۳۵، ۷۳۶، ۷۳۷، ۷۳۸، ۷۳۹، ۷۴۰، ۷۴۱، ۷۴۲، ۷۴۳، ۷۴۴، ۷۴۵، ۷۴۶، ۷۴۷، ۷۴۸، ۷۴۹، ۷۵۰، ۷۵۱، ۷۵۲، ۷۵۳، ۷۵۴، ۷۵۵، ۷۵۶، ۷۵۷، ۷۵۸، ۷۵۹، ۷۶۰، ۷۶۱، ۷۶۲، ۷۶۳، ۷۶۴، ۷۶۵، ۷۶۶، ۷۶۷، ۷۶۸، ۷۶۹، ۷۷۰، ۷۷۱، ۷۷۲، ۷۷۳، ۷۷۴، ۷۷۵، ۷۷۶، ۷۷۷، ۷۷۸، ۷۷۹، ۷۸۰، ۷۸۱، ۷۸۲، ۷۸۳، ۷۸۴، ۷۸۵، ۷۸۶، ۷۸۷، ۷۸۸، ۷۸۹، ۷۹۰، ۷۹۱، ۷۹۲، ۷۹۳، ۷۹۴، ۷۹۵، ۷۹۶، ۷۹۷، ۷۹۸، ۷۹۹، ۸۰۰، ۸۰۱، ۸۰۲، ۸۰۳، ۸۰۴، ۸۰۵، ۸۰۶، ۸۰۷، ۸۰۸، ۸۰۹، ۸۱۰، ۸۱۱، ۸۱۲، ۸۱۳، ۸۱۴، ۸۱۵، ۸۱۶، ۸۱۷، ۸۱۸، ۸۱۹، ۸۲۰، ۸۲۱، ۸۲۲، ۸۲۳، ۸۲۴، ۸۲۵، ۸۲۶، ۸۲۷، ۸۲۸، ۸۲۹، ۸۳۰، ۸۳۱، ۸۳۲، ۸۳۳، ۸۳۴، ۸۳۵، ۸۳۶، ۸۳۷، ۸۳۸، ۸۳۹، ۸۴۰، ۸۴۱، ۸۴۲، ۸۴۳، ۸۴۴، ۸۴۵، ۸۴۶، ۸۴۷، ۸۴۸، ۸۴۹، ۸۵۰، ۸۵۱، ۸۵۲، ۸۵۳، ۸۵۴، ۸۵۵، ۸۵۶، ۸۵۷، ۸۵۸، ۸۵۹، ۸۶۰، ۸۶۱، ۸۶۲، ۸۶۳، ۸۶۴، ۸۶۵، ۸۶۶، ۸۶۷، ۸۶۸، ۸۶۹، ۸۷۰، ۸۷۱، ۸۷۲، ۸۷۳، ۸۷۴، ۸۷۵، ۸۷۶، ۸۷۷، ۸۷۸، ۸۷۹، ۸۸۰، ۸۸۱، ۸۸۲، ۸۸۳، ۸۸۴، ۸۸۵، ۸۸۶، ۸۸۷، ۸۸۸، ۸۸۹، ۸۹۰، ۸۹۱، ۸۹۲، ۸۹۳، ۸۹۴، ۸۹۵، ۸۹۶، ۸۹۷، ۸۹۸، ۸۹۹، ۹۰۰، ۹۰۱، ۹۰۲، ۹۰۳، ۹۰۴، ۹۰۵، ۹۰۶، ۹۰۷، ۹۰۸، ۹۰۹، ۹۱۰، ۹۱۱، ۹۱۲، ۹۱۳، ۹۱۴، ۹۱۵، ۹۱۶، ۹۱۷، ۹۱۸، ۹۱۹، ۹۲۰، ۹۲۱، ۹۲۲، ۹۲۳، ۹۲۴، ۹۲۵، ۹۲۶، ۹۲۷، ۹۲۸، ۹۲۹، ۹۳۰، ۹۳۱، ۹۳۲، ۹۳۳، ۹۳۴، ۹۳۵، ۹۳۶، ۹۳۷، ۹۳۸، ۹۳۹، ۹۴۰، ۹۴۱، ۹۴۲، ۹۴۳، ۹۴۴، ۹۴۵، ۹۴۶، ۹۴۷، ۹۴۸، ۹۴۹، ۹۵۰، ۹۵۱، ۹۵۲، ۹۵۳، ۹۵۴، ۹۵۵، ۹۵۶، ۹۵۷، ۹۵۸، ۹۵۹، ۹۶۰، ۹۶۱، ۹۶۲، ۹۶۳، ۹۶۴، ۹۶۵، ۹۶۶، ۹۶۷، ۹۶۸، ۹۶۹، ۹۷۰، ۹۷۱، ۹۷۲، ۹۷۳، ۹۷۴، ۹۷۵، ۹۷۶، ۹۷۷، ۹۷۸، ۹۷۹، ۹۸۰، ۹۸۱، ۹۸۲، ۹۸۳، ۹۸۴، ۹۸۵، ۹۸۶، ۹۸۷، ۹۸۸، ۹۸۹، ۹۹۰، ۹۹۱، ۹۹۲، ۹۹۳، ۹۹۴، ۹۹۵، ۹۹۶، ۹۹۷، ۹۹۸، ۹۹۹، ۱۰۰۰، ۱۰۰۱، ۱۰۰۲، ۱۰۰۳، ۱۰۰۴، ۱۰۰۵، ۱۰۰۶، ۱۰۰۷، ۱۰۰۸، ۱۰۰۹، ۱۰۱۰، ۱۰۱۱، ۱۰۱۲، ۱۰۱۳، ۱۰۱۴، ۱۰۱۵، ۱۰۱۶، ۱۰۱۷، ۱۰۱۸، ۱۰۱۹، ۱۰۲۰، ۱۰۲۱، ۱۰۲۲، ۱۰۲۳، ۱۰۲۴، ۱۰۲۵، ۱۰۲۶، ۱۰۲۷، ۱۰۲۸، ۱۰۲۹، ۱۰۳۰، ۱۰۳۱، ۱۰۳۲، ۱۰۳۳، ۱۰۳۴، ۱۰۳۵، ۱۰۳۶، ۱۰۳۷، ۱۰۳۸، ۱۰۳۹، ۱۰۴۰، ۱۰۴۱، ۱۰۴۲، ۱۰۴۳، ۱۰۴۴، ۱۰۴۵، ۱۰۴۶، ۱۰۴۷، ۱۰۴۸، ۱۰۴۹، ۱۰۵۰، ۱۰۵۱، ۱۰۵۲، ۱۰۵۳، ۱۰۵۴، ۱۰۵۵، ۱۰۵۶، ۱۰۵۷، ۱۰۵۸، ۱۰۵۹، ۱۰۶۰، ۱۰۶۱، ۱۰۶۲، ۱۰۶۳، ۱۰۶۴، ۱۰۶۵، ۱۰۶۶، ۱۰۶۷، ۱۰۶۸، ۱۰۶۹، ۱۰۷۰، ۱۰۷۱، ۱۰۷۲، ۱۰۷۳، ۱۰۷۴، ۱۰۷۵، ۱۰۷۶، ۱۰۷۷، ۱۰۷۸، ۱۰۷۹، ۱۰۸۰، ۱۰۸۱، ۱۰۸۲، ۱۰۸۳، ۱۰۸۴، ۱۰۸۵، ۱۰۸۶، ۱۰۸۷، ۱۰۸۸، ۱۰۸۹، ۱۰۹۰، ۱۰۹۱، ۱۰۹۲، ۱۰۹۳، ۱۰۹۴، ۱۰۹۵، ۱۰۹۶، ۱۰۹۷، ۱۰۹۸، ۱۰۹۹، ۱۱۰۰، ۱۱۰۱، ۱۱۰۲، ۱۱۰۳، ۱۱۰۴، ۱۱۰۵، ۱۱۰۶، ۱۱۰۷، ۱۱۰۸، ۱۱۰۹، ۱۱۱۰، ۱۱۱۱، ۱۱۱۲، ۱۱۱۳، ۱۱۱۴، ۱۱۱۵، ۱۱۱۶، ۱۱۱۷، ۱۱۱۸، ۱۱۱۹، ۱۱۲۰، ۱۱۲۱، ۱۱۲۲، ۱۱۲۳، ۱۱۲۴، ۱۱۲۵، ۱۱۲۶، ۱۱۲۷، ۱۱۲۸، ۱۱۲۹، ۱۱۳۰، ۱۱۳۱، ۱۱۳۲، ۱۱۳۳، ۱۱۳۴، ۱۱۳۵، ۱۱۳۶، ۱۱۳۷، ۱۱۳۸، ۱۱۳۹، ۱۱۴۰، ۱۱۴۱، ۱۱۴۲، ۱۱۴۳، ۱۱۴۴، ۱۱۴۵، ۱۱۴۶، ۱۱۴۷، ۱۱۴۸، ۱۱۴۹، ۱۱۵۰، ۱۱۵۱، ۱۱۵۲، ۱۱۵۳، ۱۱۵۴، ۱۱۵۵، ۱۱۵۶، ۱۱۵۷، ۱۱۵۸، ۱۱۵۹، ۱۱۶۰، ۱۱۶۱، ۱۱۶۲، ۱۱۶۳، ۱۱۶۴، ۱۱۶۵، ۱۱۶۶، ۱۱۶۷، ۱۱۶۸، ۱۱۶۹، ۱۱۷۰، ۱۱۷۱، ۱۱۷۲، ۱۱۷۳، ۱۱۷۴، ۱۱۷۵، ۱۱۷۶، ۱۱۷۷، ۱۱۷۸، ۱۱۷۹، ۱۱۸۰، ۱۱۸۱، ۱۱۸۲، ۱۱۸۳، ۱۱۸۴، ۱۱۸۵، ۱۱۸۶، ۱۱۸۷، ۱۱۸۸، ۱۱۸۹، ۱۱۹۰، ۱۱۹۱، ۱۱۹۲، ۱۱۹۳، ۱۱۹۴، ۱۱۹۵، ۱۱۹۶، ۱۱۹۷، ۱۱۹۸، ۱۱۹۹، ۱۲۰۰، ۱۲۰۱، ۱۲۰۲، ۱۲۰۳، ۱۲۰۴، ۱۲۰۵، ۱۲۰۶، ۱۲۰۷، ۱۲۰۸، ۱۲۰۹، ۱۲۱۰، ۱۲۱۱، ۱۲۱۲، ۱۲۱۳، ۱۲۱۴، ۱۲۱۵، ۱۲۱۶، ۱۲۱۷، ۱۲۱۸، ۱۲۱۹، ۱۲۲۰، ۱۲۲۱، ۱۲۲۲، ۱۲۲۳، ۱۲۲۴، ۱۲۲۵، ۱۲۲۶، ۱۲۲۷، ۱۲۲۸، ۱۲۲۹، ۱۲۳۰، ۱۲۳۱، ۱۲۳۲، ۱۲۳۳، ۱۲۳۴، ۱۲۳۵، ۱۲۳۶، ۱۲۳۷، ۱۲۳۸، ۱۲۳۹، ۱۲۴۰، ۱۲۴۱، ۱۲۴۲، ۱۲۴۳، ۱۲۴۴، ۱۲۴۵، ۱۲۴۶، ۱۲۴۷، ۱۲۴۸، ۱۲۴۹، ۱۲۵۰، ۱۲۵۱، ۱۲۵۲، ۱۲۵۳، ۱۲۵۴، ۱۲۵۵، ۱۲۵۶، ۱۲۵۷، ۱۲۵۸، ۱۲۵۹، ۱۲۶۰، ۱۲۶۱، ۱۲۶۲، ۱۲۶۳، ۱۲۶۴، ۱۲۶۵، ۱۲۶۶، ۱۲۶۷، ۱۲۶۸، ۱۲۶۹، ۱۲۷۰، ۱۲۷۱، ۱۲۷۲، ۱۲۷۳، ۱۲۷۴، ۱۲۷۵، ۱۲۷۶، ۱۲۷۷، ۱۲۷۸، ۱۲۷۹، ۱۲۸۰، ۱۲۸۱، ۱۲۸۲، ۱۲۸۳، ۱۲۸۴، ۱۲۸۵، ۱۲۸۶، ۱۲۸۷، ۱۲۸۸، ۱۲۸۹، ۱۲۹۰، ۱۲۹۱، ۱۲۹۲، ۱۲۹۳، ۱

در نخستین مرحله، جمعیتی از کروموزوم‌ها به تعداد معین و بطور تصادفی تولید می‌شوند. هر کروموزوم بیانگر یک جواب از فضای جواب است و فرد نام دارد. مجموعه این افراد، جمعیت یا نسل فعلی نام دارند. کروموزوم‌ها بعد از کدگشایی تابع برازندگی بر اساس مقدار تعیین شده توسط تابع برازندگی که برابر دقت قاعده تصمیم‌گیری ایجاد شده است، تعلق می‌گیرد. از این مقدار برای هدایت فرایند انتخاب به سمت اشخاص مناسب‌تر استفاده می‌گردد. انتخاب بر اساس نخبه‌گرایی صورت می‌گیرد. تعداد جمعیت اولیه ۱۰۰، ترکیب بر اساس روش ترکیب k نقطه و جهش با احتمال ۰.۰۱ و ترکیب با احتمال ۸۰٪ انجام می‌گیرد.

در هر مرحله ۶ فرد با برازندگی بالا نسبت به کل جمعیت، انتخاب می‌شود. الگوریتم ژنتیک هنگامی دقت درخت حداقل دقت مورد علاقه خبره را تامین نماید، به پایان می‌رسد.

در نهایت از بین افراد انتخاب شده بهترین بعنوان قاعده اصلی انتخاب می‌شوند و بقیه افراد به ترتیب اندازه تابع برازندگی تابع خود در رتبه بعدی تصمیم‌گیری قرار می‌گیرد. در ابتدا خبره باید حداقل دقت درخت یا تعداد تکرار الگوریتم را مشخص کند. مزیت این روش در تعیین حداقل دقت درخت است.

۵. نتیجه‌گیری

در این تحقیق یک روش تولید درخت تصمیم بر اساس الگوریتم ژنتیک بر اساس خوشه‌بندی ارائه شده است. در این روش معیار توقف تعداد تکرار یا دقت درخت تصمیم می‌تواند باشد. ما روش بر اساس خوشه‌بندی و بدون آن را بر روی چند مجموعه داده آزمایش کردیم که نتایج آن در جدول ۱ نمایش داده شده است. شرط توقف حداقل دقت ۹۰٪ بود. الگوریتم بر روی مجموعه داده ۱ و ۲ با تعداد ۴۰ صفت و مجموعه داده‌های ۳ و ۴ با تعداد ۲۰ صفت انجام شد. برای کارهای آینده می‌توان برای درخت تصمیم برای داده‌های عددی و بر روی داده‌های واقعی ارائه داد. داده‌های آزمایش شده داده-هایی تولید شده با نرم‌افزار weka و بصورت نرمال شده تولید شده است.

جدول ۱. مقایسه سرعت دو روش بر اساس ثانیه

سرعت با داده خوشه‌بندی نشده	سرعت با داده خوشه‌بندی شده	تعداد رکورد	مجموعه داده
۱۹۸.۴	۱۱۳.۳	۲۶۶۰۰	۱
۳۷۶.۴	۲۲۱.۸	۵۲۰۰۰	۲
۴۹.۱	۲۷.۱	۶۳۰۰	۳
۳۱۲.۱	۱۸۲.۰	۳۷۰۰۰	۴

۶. مراجع

- [۱] مهدی نصیری، بهروز مینایی، مرتضی آنالویی. "تولید قواعد تصمیم با استفاده از الگوریتم ژنتیک"، دومین کنفرانس بین‌المللی تحقیق در عملیات، بابلسر، ۱۳۸۸.
- [۲] مهدی نصیری، بهروز مینایی، مرتضی آنالویی. "تولید قواعد انجمنی با استفاده از الگوریتم ژنتیک براساس خوشه بندی"، دومین کنفرانس بین‌المللی تحقیق در عملیات، بابلسر، ۱۳۸۸.
- [۳] F. Tay, And L. Shen; "A Modified Chi^2 Algorithm for Discretization", IEEE Trans. on Knowledge and Data Eng., vol. ۱۴, no. ۳, pp. ۶۶۶-۶۷۰, ۲۰۰۲.
- [۴] T. M. Mitchel; "Machine Learning", McGraw-Hill, ۱۹۹۷.
- [۵] M. Chemielewski and J. Grzrmala-Busse; "Global Discretization of Continues Attributes as Preprocessing for Machine Learning", Int. J. of Approximate Reasoning, vol. ۱۵, no. ۴, pp. ۳۱۹-۳۳۱, ۱۹۹۶.
- [۶] J. Dougherty, R. Kohavi, and M. Sahami; "Supervised and Unsupervised Discretization of Continuous Features", in Proc. of ۱۲th Int. Conf. on Machine Learning, pp. ۱۹۴-۲۰۲, ۱۹۹۵.
- [۷] U.M. Fayyad and K.B. Irani; "Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning", in Proc. of ۱۳th Int. Joint Conf. on Artificial Intelligence, pp. ۱۰۲۲-۱۰۲۷, ۱۹۹۳.
- [۸] J. R. Quinlan; "Induction on Decision Trees", Machine Learning, vol. ۱, pp. ۸۱-۱۰۶, ۱۹۸۶.
- [۹] I. Breiman, J. H. Friedman, R. A. Oslen, and C. J. Stone; "Classification and Regression Trees", Belmont, CA: Wadsworth, ۱۹۸۴.
- [۱۰] J. R. Quinlan; "C۴.۵: Programs for Machine Learning", San Mateo, CA: Morgan Kaufmann", ۱۹۹۳.

- [11] L. Hyafil and R.L. Rivest; “*Constructing Optimal Binary Decision Trees is NP-Complete*”, Information Processing Letters, vol. 5, no. 1, pp. 15–17, 1976.
- [12] J. Mingers; “*An Empirical Comparison of Selection Measures for Decision Tree Induction*”, Machine Learning, vol. 2, pp. 319–342, 1989.
- [13] L.A. Zadeh; “*Fuzzy Sets, Information Control*”, vol. 8, pp. 338–353, 1965.
- [14] Li-XinWang; “*A Course in Fuzzy Systems and Control*”, Prentice Hall, 1997.
- [15] Q.R.Wang, and C.Y.Suen; “*Large Tree Classifier with Heuristic Search and Global Training*”, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 9, no. 1, pp. 91–102, 1987.
- [16] P.L.Carbone; “*Data Mining or Knowledge Discovery in Databases: An Overview*”, © MITRE Corporation, 1997
- [17] U.M.Fayyad, G.Piatetsky-Shaprio and P.Smyth; “*From Data Mining to Knowledge Discovery: an Overview*” Cambridge, MA: AAAI/MIT, pp. 1–34, 1996.
- [18] G. Hulten, L. Spencer, P. Domingos; “*Mining time-changing data streams*”, In Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD), pages 97–106, 2001.
- [19] H. Wang, W. Fan, P. S. Yu, J. Han; “*Mining concept drifting data streams using ensemble classifiers*”, In Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD), pages 226–235, 2003.
- [20] W. Cohen; “*Fast effective rule induction*”, In Proceedings of the 12th International Conference on Machine Learning (ICML), pages 115–123, 1995
- [21] T. G. Dietterich; “*An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization*”, Machine Learning, 40:139–158, 2000.
- [22] Pang-Ning Tan, Michael Steinbach, Vipin Kumar; “*Introduction to Data Mining*”, Addison Wesley, 2006.
- [23] A.E. Eiben and J.E. Smith; “*Introduction to Evolutionary Computing*”, Springer, 2003.
- [24] Mehmed Kantardzic; “*DATA MINING - Concepts, Models, Methods, and Algorithms*”, John Wiley & Sons, 2003.

¹ Attribute

² Class Label

³ Comprehensibility

⁴ Inductive Learning

⁵ Discrete-Valued Functions

⁶ Discretization

⁷ Partitioning

⁸ Covering

⁹ Supervised

¹⁰ Unsupervised

¹¹ Recursive Splitting

¹² Data-Driven

¹³ Branching Attribute

¹⁴ Overfitting

¹⁵ Pre-Pruning

¹⁶ Post-Pruning

¹⁷ Genetic Algorithm (GA)

¹⁸ Fitness

¹⁹ Offspring

²⁰ John Holland