

یادگیری درخت تصمیم فازی

محمد رضا گودرزی، محمد رضا کنگاوری

دانشگاه علم و صنعت ایران، دانشکده مهندسی کامپیوتر
mrgoodarzi@mail.iust.ac.ir

چکیده

یادگیری استقرایی به وسیله درخت‌های تصمیم یکی از متداول‌ترین روش‌های اخذ دانش می‌باشد. دانش حاصل از به کارگیری این روش، دارای این مزیت مهم می‌باشد که برای انسان از قابلیت درک خوبی برخوردار است. با گسترش استفاده از نمایش فازی، محققان تلاش کرده‌اند دو رهیافت درخت‌های تصمیم نمادین و منطق فازی را با یکدیگر ترکیب کنند تا ضمن حفظ قابلیت درک، انعطاف‌پذیری بیشتری در یادگیری مفاهیم دنیای واقعی و نیز مقابله با اغتشاش حاصل شود. در این مقاله، به بررسی نحوه یادگیری درخت‌های تصمیم فازی و برخی کاربردهای آن پرداخته شده است.

واژه‌های کلیدی: یادگیری ماشین-درخت‌های تصمیم-یادگیری فازی-درخت‌های تصمیم فازی

مقدمه

اخذ دانش (یادگیری) و تفسیر آن (استدلال) دو بخش اصلی در یک سیستم مبتنی بر دانش محسوب می‌شوند [1]. با افزایش میزان مقادیر داده، اخذ اتوماتیک دانش (Automatic Knowledge Acquisition) روز به روز از اهمیت بیشتری برخوردار می‌شود. هنگامی که هر عنصر داده‌ای دارای یک کلاس مشخص بوده، دانش زمینه به صورت گسترده‌ای وجود نداشته، و هدف، دستیابی به دانشی باشد که قادر به توصیف این کلاس‌ها باشد (چه برای استدلال در مورد آنها و چه برای کلاس بندی داده‌های آینده)، این فرایند اکتساب دانش، یادگیری تحت نظارت از روی نمونه‌ها نامیده می‌شود. از آنجا که هدف از این یادگیری استنتاج دانش جدید می‌باشد، لذا می‌بایست یک فرایند استقرا (Induction) برای این منظور به خدمت گرفته شود. هنگامی که هم زبان توصیف داده‌های آموزشی و هم زبان توصیف دانش منتجه، خصیصه‌های نمادین را مورد استفاده قرار دهند، این فرایند یادگیری را یادگیری نمادین (Symbolic Learning) می‌نامیم [2].

یکی از متداول‌ترین روش‌های یادگیری نمادین، استقرای درخت تصمیم می‌باشد که نخستین بار به وسیله کوئینلان (Quinlan) تحت نام الگوریتم $ID3$ توسعه داده شد [2]. این روش در ابتدا برای داده‌های نمادین، آن هم زمانی که تمامی اطلاعات مورد نیاز مهیا بوده باشند، توسعه داده شد. دانش به دست آمده، با یک درخت تصمیم نمادین بازنمایی می‌شود که برای انسان از قابلیت درک خوبی برخوردار است. یک رویه استنتاجی ساده در کنار این ساختار درختی قرار می‌گیرد که انتساب تصمیمات نمادین به داده‌های جدید را امکان‌پذیر می‌سازد (یعنی انتساب کلاس به نمونه‌ها). درختان تصمیم نمادین به سهولت قابلیت تبدیل به مجموعه‌ای از قوانین را دارا بوده، لذا می‌توانند از آن پس در یک سیستم مبتنی بر قوانین به کار گرفته شوند [3].

در کاربردهای واقعی، داده‌ها به ندرت دارای این قابلیت هستند که بتوانند در محدودیت‌های اعمال شده از سوی چنین الگوریتمی جا داده شوند. چنین محدودیتهایی به صورت مختلفی بروز پیدا می‌کنند [4]. نخست اینکه، یادگیری استقرایی

نمادین، تنها بر روی قلمروهای نمادین قابل به‌کارگیری است، درحالی که برخی یا تمامی خصیصه‌ها می‌توانند به صورت چندمقداره و یا پیوسته توصیف داده شوند. همچنین ممکن است مقادیر برخی از خصیصه‌ها معلوم نباشد که این می‌تواند هم در داده‌های آموزشی و هم در زمان تصمیم‌گیری اتفاق بیفتد. این درحالی است که در هنگام تصمیم‌گیری، می‌بایست بر اساس هر آنچه اطلاعات که فراهم است، تصمیم اخذ شود (و یا یک تست جدید پیشنهاد گردد). علاوه بر این، داده‌ها می‌توانند مغشوش (noisy) یا اشتباه (erroneous) باشند. اگر چه مورد دوم می‌تواند به حداقل رسانده شود، ولیکن مورد نخست، به ویژه در قلمروهای پیوسته، می‌بایست مورد توجه قرار گیرد. سرانجام اینکه خصیصه‌ها ممکن است دربرگیرنده ترم‌های زبانی باشند که تعاریف دقیق و غیرمبهمی برای آنها وجود نداشته باشد.

این قبیل مسائل در مورد درخت‌های تصمیم مورد کنکاش قرار گرفته، بهبودهایی را در زمینه روش‌های ساخت آنها موجب گردیده‌اند. به عنوان مثال، برای رفتار با داده‌های پیوسته و نیز قابلیت به کارگیری درخت در زمینه رگرسیون، روش CART پیشنهاد شده است [5]. با این حال، درخت‌های ساخته شده با استفاده از این روش، از قابلیت درک پایین‌تری برای انسان برخوردار می‌باشند که بسیاری از اوقات خوشایند نمی‌باشد. در درخت‌های تصمیم نمادینی که درگیر قلمروهای پیوسته یا چندمقداره هستند، پیشنهاد شده است که بلوک‌های افزایش این قلمروها به عنوان مقادیر جدید برای چنین خصیصه‌هایی به کارگرفته شوند. در مواردی که چنین بلوک‌هایی هم‌پوشانی داشته باشند، یک رویه استنتاج احتمالاتی می‌تواند مورد استفاده قرار گیرد. روش‌هایی نیز برای رفتار با خصیصه‌های نامعلوم و نیز مغشوش، مورد مطالعه قرار گرفته‌اند [3]، [6] و [7].

در سال‌های اخیر، استفاده از بازنمایی فازی (Fuzzy Representation) به واسطه وجود قابلیت‌های خوب آن، عمومیت زیادی پیدا نموده است. این نوع بازنمایی که مبتنی بر مجموعه‌های فازی و مورد استفاده استدلال تقریبی (Approximate Reasoning) می‌باشد، به ویژه، برای پرکردن خلا مفهومی میان خصیصه‌های ذهنی/مبهم با داده‌های کمی قابل به‌کارگیری می‌باشد [8] و [9]. به واسطه تغییر ذره‌ای مرزهای مجموعه‌های فازی و امکان استدلال تقریبی و پردازش تدریجی، نمایش فازی نیز برای رفتار با داده‌های غیردقیق و مغشوش مناسب می‌باشد.

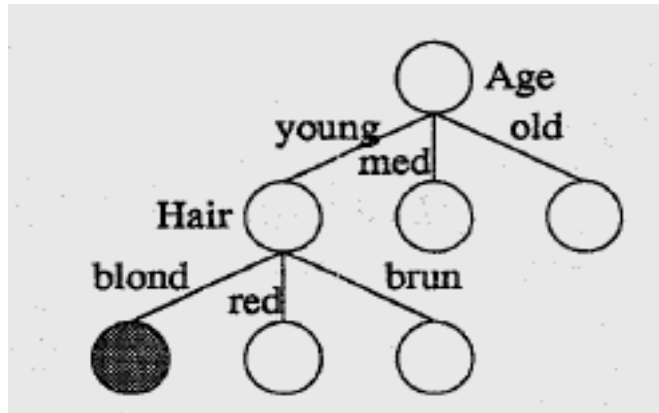
لذا در دهه اخیر، برخی از محققان تلاش کرده‌اند منطق فازی به همراه قابلیت‌های استدلال تقریبی آن با الگوریتم‌های استقرای درخت تصمیم نمادین در یکدیگر ادغام شوند. هدف از چنین ادغامی، ایجاد درخت‌های تصمیم فازی (Fuzzy Decision Trees) می‌باشد که می‌توانند از یک سو قابلیت درک بالا، و از سوی دیگر قابلیت رفتار با قلمروهای پیوسته، ترم‌های زبانی فازی، و اندازه‌گیری‌های مغشوش را داشته باشند. اولی از ویژگی‌های درخت‌های تصمیم و دومی از خصوصیات سیستم‌های فازی ناشی می‌شوند. از آنجا که در درخت‌های تصمیم روش‌هایی برای رفتار با خصیصه‌های نامعلوم وجود دارند، چنین روش‌هایی می‌توانند به سهولت در درخت‌های تصمیم فازی نیز به کارگرفته شوند، و از آنجا که درخت‌های تصمیم می‌توانند به صورت پایگاه‌های قوانین تفسیر شوند، درخت‌های تصمیم فازی نیز می‌توانند به عنوان وسیله‌ای برای یادگیری قوانین فازی مدنظر قرار گیرند.

این مقاله، با هدف معرفی درخت‌های تصمیم فازی و ارائه برخی از کاربرهای آنها نگاشته شده است. از این رو، در قسمت‌های بعدی این مقاله، ابتدا روش ID3 برای استقرای درخت تصمیم نمادین توضیح داده می‌شود. آنگاه روش استقرای درخت تصمیم فازی که تغییر یافته روش ID3 می‌باشد، شرح داده شده، و سپس مروری بر چند کاربرد این نوع درخت‌ها صورت خواهد گرفت.

درخت‌های تصمیم نمادین

در الگوریتم‌های استقرای درخت تصمیم، فرآیند ساخت درخت به وسیله نمونه‌هایی به پیش برده می‌شود که به وسیله زوج‌های «خصیصه-ارزش» (Attribute-Value Pairs) و یک برچسب کلاس (Class Label) توصیف داده شده‌اند. هر شاخه از چنین درختی با یک شرط «مقدار= V » برچسب‌گذاری می‌شود. برای رسیدن به یک گره برگ، تمامی شرایط روی مسیر آن

می بایست برآورده شوند. به منظور استنتاج برای یک نمونه جدید، مقادیر خصیصه‌های آن با این شرایط مطابقت داده شده، و نمونه براساس برجسب کلاس داده‌های آموزشی یافت شده در گره برگ ارضا شده، کلاس‌بندی می‌شود. شکل (۱)، یک درخت تصمیم نمونه را نشان می‌دهد.



شکل ۱- یک درخت تصمیم نمونه

ساخت درخت مبتنی بر رویه افراز بازگشتی (Recursive Partitioning) می‌باشد. برای سهولت محاسباتی، همواره خصیصه‌ها مستقل از یکدیگر فرض می‌شوند. $ID3$ و $CART$ دو روش متداول برای یادگیری چنین درختانی می‌باشند. در حالی که $ID3$ قابلیت فهم دانش را هدف خود قرار داده و در قلمروهای نمادین قابل به‌کارگیری است، $CART$ به طور طبیعی برای رفتار با قلمروهای پیوسته طراحی شده ولیکن دچار مشکل کمبود قابلیت درک می‌باشد.

رویه افراز بازگشتی در هر زمان یک خصیصه را برای بسط گره انتخاب می‌کند. معمولاً خصیصه‌ای انتخاب می‌شود که یک معیار اطلاعاتی را برای نمونه‌های آموزشی ارضاکننده شرایط منتهی به این گره بیشینه کند. با استفاده از مقادیر قلمروی چنین خصیصه‌ای، نمونه‌های موجود در این گره بین گره‌های فرزند تقسیم می‌شوند. به این ترتیب که به هر گره فرزند نمونه‌هایی راه می‌یابند که یک شرط اضافی منتهی به آن گره را ارضا نموده باشند. آنگاه، همین رویه به طور بازگشتی برای هر گره فرزند تکرار می‌شود، مگر هنگامی که تمامی خصیصه‌ها روی مسیر از گره ریشه تا آن گره مورد استفاده قرار گرفته باشند، یا هنگامی که تمامی نمونه‌های موجود در این گره متعلق به یک کلاس واحد بوده باشند. برخی محک‌های دیگر نیز ممکن است مدنظر قرار گیرند. به عنوان مثال، برای درخت نشان داده شده در شکل (۱)، اگر نمونه‌های موجود در گره سیاه هنوز هم دارای کلاس‌بندی‌های ناهمگون بوده و در عین حال خصیصه(های) دیگری برای بسط این گره موجود بوده باشند، این امکان وجود دارد که این گره گسترش بیشتری پیدا نماید. فرآیند بازگشتی ساخت درخت می‌تواند به صورت زیر شرح داده شود:

۱- برای هر گره N محتوای اطلاعاتی را از رابطه زیر محاسبه کن:

$$I_N = -\sum_{k=1}^{|C|} p_k \cdot \log_2(p_k) \quad (1)$$

که در آن C مجموعه تصمیمات (کلاس‌ها) و p_k احتمال تعلق یک نمونه موجود در این گره به کلاس k ام می‌باشد.

۲- برای هر خصیصه باقیمانده a_i (یعنی خصیصه‌ای که قبلاً روی مسیر از گره ریشه تا این گره به کار گرفته نشده باشد)، بهره اطلاعاتی حاصل از تقسیم این گره با استفاده از این خصیصه را با به‌کارگیری رابطه زیر محاسبه کن:

$$G_i = I_N - \sum_{j=1}^{|D_i|} w_j \cdot I_{N_j} \quad (2)$$

که در آن D_i بیانگر مجموعه ویژگی‌های متناظر با a_i ، I_{N_j} محتوای اطلاعاتی در فرزند j ام N ، و w_j نسبت نمونه‌هایی از گره N می‌باشد که شرایط منتهی به گره N_j را ارضا نموده‌اند.

۳- این گره را با استفاده از خصیصه‌ای که بهره را بیشینه می‌کند، بسط بده.

رویه فوق برای ساخت درخت، در واقع یک افراز از فضای توصیف را ایجاد می‌کند، با این هدف که «بلوک‌های بزرگی» از این فضای توصیف حاصل شوند، به طوری که داده‌های آموزشی در هر یک از این بلوک‌ها دارای کلاس‌بندی‌های یکسان باشند. کاملاً طبیعی به نظر می‌رسد که این کلاس‌بندی‌های واحد را مبنای تصمیم‌گیری برای کلاس یک نمونه جدید قرار دهیم. به این ترتیب که یک عنصر داده‌ای جدید به همان صورتی کلاس‌بندی می‌شود که داده‌های آموزشی واقع در همان بلوک. البته، اگر در بلوکی (بلوک‌هایی) از این افراز، داده‌های آموزشی دارای کلاس‌های متفاوت وجود داشته باشند، مشکلاتی بروز خواهند نمود. این امر به چندین دلیل ممکن است رخ دهد. مثلاً به واسطه وجود مجموعه‌ای ناکافی از خصیصه‌ها، اغتشاش و یا خطا. مسأله بالقوه دیگری که می‌تواند به وجود آید، عبارت از این می‌باشد که بلوکی وجود داشته باشد که دارای هیچ داده آموزشی نباشد. این نیز می‌تواند به این دلیل رخ دهد که مجموعه آموزشی دارای تعداد ناکافی از نمونه‌ها بوده باشد. با پیروی از رویه تصمیم‌گیری ذکر شده در بالا، در مرحله استنتاج، خصیصه‌های یک نمونه جدید، در برابر شرایط حاضر در درخت مقایسه می‌شوند. آشکار است چنین اقدامی متناظر با تعیین بلوک افرازی می‌باشد که این نمونه در آن قرار می‌گیرد. کلاس نمونه‌های گره برگ که شرایط آن به وسیله این نمونه برآورده شده‌اند، به عنوان کلاس منسوب به این نمونه بازگردانده می‌شود. به عنوان مثال، اگر فرض شود که گره سیاه در شکل (۱) دارای نمونه‌هایی با یک کلاس یکسان است، هر نمونه دیگری که با همان دو شرط «سن=جوان» و «مو=بلوند» توصیف شود، نیز متعلق به همان کلاسی در نظر گرفته می‌شود که نمونه‌های این گره به آن تعلق دارند.

روش $ID3$ چنین فرض می‌کند که تمامی خصیصه‌ها نمادین می‌باشند. گسترش‌هایی از این الگوریتم ارائه شده‌اند (مثلاً $C4.5$ [3]) که قابلیت رفتار با خصیصه‌های نامعلوم (Unknown)، همچنین ناسازگاری (Inconsistency) درخت، و یا نقصان (Incompleteness) آن را دارا می‌باشند. مشکل ناسازگاری در درخت تصمیم زمانی رخ می‌دهد که یک گره برگ دارای نمونه‌هایی متعلق به بیش از یک کلاس بوده باشد. همچنین زمانی که هیچ یک از شرایط خارج شونده از گره برآورده نشوند، هیچ گره برگ ارضا نشده، مشکل نقصان به وجود می‌آید.

درخت‌های تصمیم فازی

در درخت‌های تصمیم فازی رسیدن به میزان بالایی از قابلیت درک از یک سو، و امکان رفتار تدریجی از سوی دیگر به عنوان دو هدف اصلی مطرح می‌باشند. روش‌های مختلفی برای توسعه درخت‌های تصمیم فازی ارائه شده‌اند ([10]-[19]). روشی که در اینجا مورد بررسی قرار می‌دهیم، الگوریتمی است که گسترش یافته روش سنتی $ID3$ می‌باشد [14]. برای یادگیری درخت‌های تصمیم فازی، ابتدا لازم است نمونه‌ها دارای نمایش فازی باشند. بازنمایی دقیق به عنوان حالت خاصی از بازنمایی فازی در نظر گرفته می‌شود. به طور کلی، نمایش فازی یک نمونه نیز می‌تواند به صورت مجموعه‌ای از زوج‌های «خصیصه=ارزش» در نظر گرفته شود، ولی در اینجا ارزش خصیصه یک مجموعه فازی می‌باشد. مطابق با نوع مجموعه فازی، نمایش فازی ارزش خصیصه می‌تواند یکی از انواع مقادیر زبانی، مقادیر حقیقی، مقادیر بازه‌ای، اعداد فازی، بردار فازی، و یا ترکیبی از اینها در نظر گرفته شود.

همچنین در حالی که در درخت‌های تصمیم نمادین کلاس نمونه‌ها دقیق می‌باشد (غالباً یک کلاس مثبت و یک کلاس منفی) در درخت‌های تصمیم فازی فرض می‌شود که کلاس‌ها نیز دارای یک نمایش فازی هستند؛ یعنی کلاس یک نمونه، دیگر الزاماً یک کلاس دقیق معین نمی‌باشد، بلکه می‌تواند مثلاً یک بردار فازی باشد که روی فضای کلاس‌ها تعریف شده است. به عنوان مثال زمانی که تنها دو کلاس P و N وجود دارند، نمایش فازی کلاس نمونه می‌تواند برداری همانند $(0.4/P, 0.6/N)$ بوده باشد.

یک درخت تصمیم فازی (FDT) تعمیم یک درخت تصمیم دقیق (CDT) می‌باشد. تفاوت‌های اصلی میان CDT و FDT عبارتند از [20]:

- ۱- FDT یک افراز فازی از فضای توصیف X را ارائه می‌کند درحالی که CDT یک افراز دقیق از آن را ارائه می‌نماید.
- ۲- گره FDT یک مجموعه فازی است که بر روی X تعریف شده است در حالی که هر گره CDT یک مجموعه دقیق از X می‌باشد.
- ۳- اشتراک گره‌های قرارگرفته بر روی یک سطح در FDT غیر تهی ولی در CDT تهی می‌باشد.
- ۴- در حالت فازی، اگر N یک گره داخلی و $\{N_i\}$ مجموعه تمامی فرزندان N باشد، آنگاه $N_i \subset N$ ولی در حالت دقیق، تساوی $N_i = N$ برقرار می‌باشد.
- ۵- در حالت فازی هر مقدار خصیصه به عنوان یک مجموعه فازی ولی در حالت دقیق هر مقدار خصیصه به عنوان یک مجموعه دقیق در نظر گرفته می‌شود.
- ۶- در FDT هر مسیر از ریشه تا برگ می‌تواند به یک قانون فازی به همراه یک درجه درستی تبدیل شود ولی در CDT به یک قانون دقیق تبدیل می‌شود.
- ۷- در CDT نمونه‌ای که قرار است کلاس‌بندی شود، تنها با یک مسیر مطابقت می‌یابد ولی در FDT می‌تواند با چندین مسیر مطابقت پیدا نماید.

رویه ساخت درخت فازی از همان رویه ساخت $ID3$ پیروی می‌کند، با این تفاوت که بهره اطلاعاتی خصیصه‌های منفرد در اینجا با استفاده از مجموعه‌های فازی، درجات عضویت، و روش‌های استدلال ارزیابی می‌گردد [4]. به طور مشخص‌تر، تفاوت اصلی در نحوه برآورد احتمالات p_k می‌باشد. در درخت‌های تصمیم فازی، درجه تعلق یک نمونه در یک گره یک عدد حقیقی در بازه $[0,1]$ بوده، بیانگر میزان ارضای محدودیت‌های فازی منتهی به آن گره می‌باشد. فرض کنید عضویت نمونه e در گره N بوده باشد. به وضوح $\mu_e^{Root} = 1$ ، و نیز $\mu_e^{N_j} = t(\mu_e^N, \mu_e^{A_j})$ که در آن N_j عبارتست از j امین فرزند N ، و A_j ترم فازی متناظر با محدودیت فازی منتهی به N_j می‌باشد. همچنین عملگر $t(\cdot, \cdot)$ بیانگر یک t -نرم می‌باشد. در این صورت، در گره N تعداد رخدادهای برای کلاس زبانی فازی k برابر است با: $\sum_{e \in E} t(\mu_e^N, \mu_e^{A_k})$ که E مجموعه کلیه نمونه‌ها می‌باشد. اکنون، احتمال p_k می‌تواند با استفاده از رابطه زیر:

$$p_k = \frac{\sum_{e \in E} t(\mu_e^N, \mu_e^{A_k})}{\sum_{k' \in C} \sum_{e \in E} t(\mu_e^N, \mu_e^{A_{k'}})} \quad (3)$$

تخمین زده شود. با داشتن احتمالات p_k محتوای اطلاعاتی مطابق رابطه (۱) محاسبه شده، درخت می‌تواند مشابه با روشی که پیش‌تر ذکر شد، گسترش پیدا نماید.

همچنین در زمان استنتاج، گام اول عبارتست از بررسی اینکه یک نمونه جدید با چه درجه‌ای هر یک از گره‌های برگ را برآورده می‌سازد. به عبارت دیگر ضروری است درجه تعلق آن در هر یک از بلوک‌های افراز فازی حاصل یافت شود [4]. با استفاده از همان رویه‌ای که تعلق یک نمونه آموزشی در یک گره درخت را تعیین می‌کرد، یعنی، تعیین درجه تعلق نمونه در محدودیت‌های فازی انشعاب‌ها، و ترکیب آنها با استفاده از یک عملگر t -نرم، می‌توان درجه تعلق یک نمونه جدید در هر گره برگ را محاسبه نمود. در مرحله بعد لازم است برای یک نمونه رسیده به یک گره برگ به خصوص، در مورد کلاس آن تصمیم‌گیری شود. در درخت‌های تصمیم فازی بسیار محتمل است در گره‌های برگ نمونه‌هایی متعلق به کلاس‌های متفاوت وجود داشته باشند. به ویژه، در مواقعی که برجسب کلاس یک ترم زبانی فازی (یا سخت) نمی‌باشد، بلکه مقداری از قلمروی کلاس است (مثلاً وقتی که این قلمرو پیوسته باشد)، آنگاه هیچ‌گونه کلاس‌بندی یکسانی برای نمونه‌های گره برگ وجود نداشته، و تنها می‌توان درباره درجه تعلق یک نمونه، در کلاس‌های زبانی، سخن گفت. چنین اطلاعات کلاسی را می‌توان به روش‌های مختلفی تفسیر نمود. برای مثال، می‌توان آن را به عنوان اجتماعی از مجموعه‌های فازی (دارای یک ضریب وزنی) و یا

یک ترکیب فصلی از گزاره‌های فازی مربوط به انتساب کلاس (وزن دار شده با نسبت نمونه‌های آن گره برگ که در مجموعه‌های فازی متناظر با کلاس‌های زبانی موجود هستند) در نظر گرفت.

چندین گره برگ که با درجات مختلفی به وسیله یک نمونه جدید ارضا می‌شوند، منجر به ایجاد نوع دومی از تضاد (Conflict) می‌گردند. این تضادها از ویژگی همپوشانی مجموعه‌های فازی ناشی می‌شوند. فقدان خصیصه‌های نامعلوم در این نمونه، و دانش ناقص (فقدان انشعابی که نمونه بتواند آن را برآورده نماید) نیز از دلایل دیگر این پدیده محسوب می‌گردند. از آنجا که هر گره برگ می‌تواند به عنوان یک قانون فازی تفسیر شود، آنگاه برگ‌های چندگانه به عنوان ترکیبات فصلی قوانین فازی قابل تفسیر می‌باشند.

به دلیل تنوع تفسیرهای مختلفی که می‌توان درباره ارضای گره‌های مختلف و شیوه ترکیب آن در نظر گرفت، برای درخت‌های تصمیم فازی استنتاج‌های گوناگونی مطرح شده‌اند [4] و [14]-[16]. این استنتاج‌ها یا از اصول استدلال تقریبی ناشی می‌شوند و یا مبتنی بر استنتاج در یادگیری نمونه‌ای می‌باشند. در اینجا تنها به یکی از این موارد، زمانی که کلاس نمونه‌ها ترم‌های زبانی (فازی یا سخت) هستند، اشاره می‌نماییم ([15] و [16]).

ابتدا پس از ساخته شدن درخت، برای هر گره برگ یک برچسب کلاسی تعیین می‌شود. مثلاً می‌توان کلاسی که بیشترین نمونه را داراست، و یا نسبت نمونه‌های آن به کل نمونه‌های موجود در گره از یک آستانه بیشتر است، به عنوان برچسب کلاس در نظر گرفته، و از دیگر داده‌های موجود در گره صرف‌نظر نمود. روش بهتر اینست که همه داده‌ها را به حساب آورده و مجموعه تمامی کلاس‌ها به همراه درجه تعلق مربوطه به آنها نسبت داده شود، به این صورت که:

$$CL(N) = \{(c, \kappa_N(c)) | c \in C\} \quad (4)$$

در رابطه فوق، $\kappa_N(c)$ درجه بازنمایی نرمال شده کلاس c مربوط به گره N می‌باشد؛ یعنی:

$$\kappa_N(c) = \frac{\delta_N^c}{\sum_j \delta_N^j} \quad (5)$$

که δ_N^c مجموع درجات عضویت نمونه‌های متعلق به کلاس c در گره N می‌باشد؛ یعنی:

$$\delta_N^c = \sum_{e \text{ class} = c, \mu_e^N > 0} \mu_e^N \quad (6)$$

به عنوان مثال، داده‌های آموزشی نشان داده شده در شکل (۲-الف) و مجموعه‌های فازی خصیصه‌ها، نشان داده شده در شکل (۲-ب) را در نظر بگیرید [15]. یک نکته جالب در اینجا مشاهده می‌شود و آن اینست که داده‌های سطرهای چهارم و ششم جدول به وضوح ناسازگار می‌باشند. ابتدا نحوه تولید درخت تصمیم فازی را به صورتی مختصر شرح می‌دهیم.

	μ	height	weight	hair color	class
low = {1/160, 0.8/165, 0.5/170, 0.2/175}	1	160	60	blond	C_1
middle = {0.5/165, 1/170, 0.5/175}	0.8	180	80	black	C_2
high = {0.2/165, 0.5/170, 0.8/175, 1/180}	0.2	170	75	black	C_2
light = {1/60, 0.8/65, 0.5/70, 0.2/75}	0.7	175	60	red	C_1
middle = {0.5/65, 1/70, 0.5/75}	1	160	75	black	C_2
heavy = {0.2/65, 0.5/70, 0.8/75, 1/80}	0.3	175	60	red	C_2
light = {1/blond, 0.3/red}	1	165	60	blond	C_2
dark = {0.6/red, 1/black}	0.5	180	70	blond	C_1

(ب)

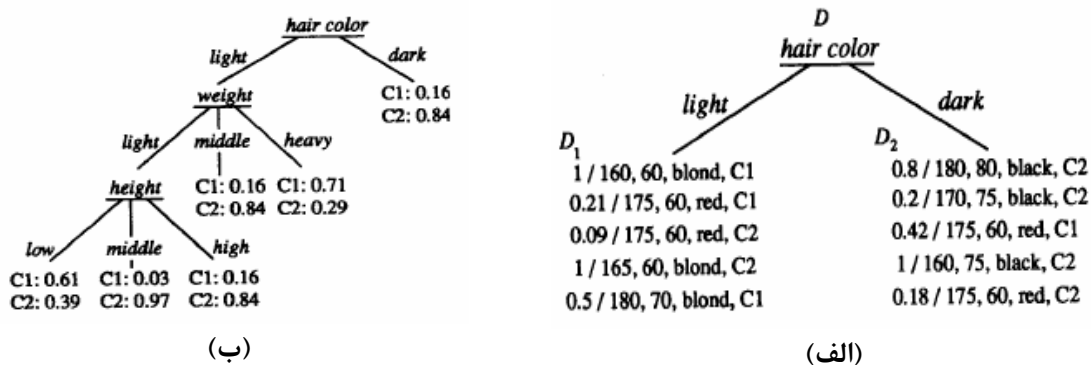
(الف)

شکل ۲- الف) داده‌های آموزشی (ب) مجموعه‌های فازی مربوط به خصیصه‌ها

برای گره ریشه، بهره اطلاعاتی ناشی از تقسیم این گره با استفاده از خصیصه «قد» به این صورت محاسبه می‌شود. در این گره محتوای اطلاعاتی برابر می‌باشد با: $I_{Root} = -\frac{2.2}{5.5} \log_2 \frac{2.2}{5.5} - \frac{3.3}{5.5} \log_2 \frac{3.3}{5.5} = 0.971(bits)$. برای خصیصه «قد» با توجه به وجود سه مجموعه فازی «کوتاه»، «متوسط» و «بلند» برای این خصیصه، ابتدا برای گره فرزند حاصل از مجموعه فازی «کوتاه» خواهیم داشت: $I_{Root,height,low} = -\frac{1.14}{3.1} \log_2 \frac{1.14}{3.1} - \frac{1.96}{3.1} \log_2 \frac{1.96}{3.1} = 0.949(bits)$. به همین ترتیب، $I_{Root,height,middle} = 0.871(bits)$ و $I_{Root,height,high} = 0.990(bits)$. لذا بهره اطلاعاتی در صورتی که این خصیصه به عنوان خصیصه تست در این گره انتخاب شود، برابر خواهد بود با:

$$G_{height} = 0.971 - \left(\frac{3.1}{6.7} \times 0.949 + \frac{1.2}{6.7} \times 0.871 + \frac{2.4}{6.7} \times 0.990 \right) = 0.021(bits)$$

محاسبات مشابه برای دو خصیصه دیگر نتیجه خواهند داد: $G_{weight} = 0.118(bits)$ و $G_{hair\ color} = 0.164(bits)$. با توجه به اینکه بهره اطلاعاتی به دست آمده به ازای استفاده از خصیصه «رنگ مو» بیش از بهره اطلاعاتی به دست آمده از دو خصیصه دیگر می‌باشد، این خصیصه به عنوان خصیصه تست در گره ریشه انتخاب می‌شود. در نتیجه نمونه‌های موجود در گره ریشه به صورت نشان داده شده در شکل (۳-الف) در بین دو گره فرزند حاصل تقسیم می‌شوند. با اجرای الگوریتم به صورت مشابه، درخت نشان داده شده در شکل (۳-ب) حاصل خواهد گردید. توجه شود که در اینجا به جای گسترش حداکثری درخت، آن را تا جایی گسترش داده‌ایم که یا خصیصه دیگری باقی نمانده باشد و یا «بیشتر» نمونه‌های یک گره متعلق به یک کلاس واحد بوده باشند. همچنانکه دیده می‌شود، برگ‌های این درخت، پس از ساخته شدن آن، به وسیله یک درجه بازنمایی نرمال شده مربوط به کلاس‌های مختلف برچسب‌گذاری گردیده‌اند.

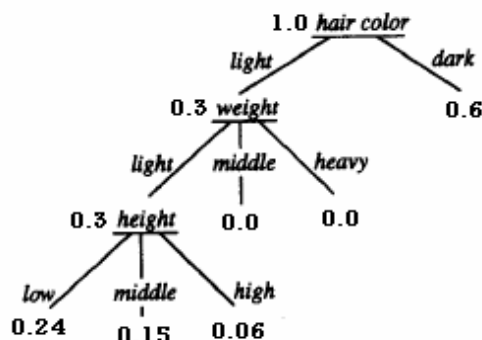


شکل ۴- (الف) زیردرخت حاصل از بسط گره ریشه با استفاده از خصیصه «رنگ مو» (ب) درخت کامل شده

در هنگام استنتاج از روی این درخت فازی، پس از تعیین درجه تعلق نمونه در هر یک از گره‌های برگ، می‌توان درجه تعلق نمونه به هر یک از کلاس‌ها را به صورت زیر محاسبه کرد:

$$\mu_e^c = \sum_{N \in L} \mu_e^N \cdot \kappa_N(c) \quad (7)$$

که در آن L مجموعه گره‌های برگ می‌باشد. مثلاً در درخت نشان داده شده در شکل (۴-ب)، برای نمونه جدیدی با ویژگی‌های «قد=۱۶۵»، «وزن=۶۰» و «رنگ مو=قرمز»، گره‌های مختلف برگ به صورتی ارضا می‌شوند که در شکل (۵) نشان داده شده است.



شکل ۵- درجهٔ ارضای گره‌های گوناگون درخت برای یک نمونهٔ به خصوص

اکنون برای این نمونه

$$\mu_e^{C_1} = 0.24 \times 0.61 + 0.15 \times 0.03 + 0.06 \times 0.16 + 0.6 \times 0.16 = 0.25 \quad (8)$$

$$\mu_e^{C_2} = +0.24 \times 0.39 + 0.15 \times 0.97 + 0.06 \times 0.84 + 0.6 \times 0.86 = 0.79 \quad (9)$$

لذا اگر قرار است به این نمونه در نهایت یک کلاس منسوب شود، کلاس C_2 منسوب خواهد گردید.

کاربردهای درخت‌های تصمیم فازی

به دلیل مزایای مهم درخت‌های تصمیم فازی، کاربردهای بسیاری برای آنها گزارش شده‌اند. به عنوان یک مثال جالب، [21] درخت‌های تصمیم فازی را برای حل مسألهٔ تصدیق هویت با استفاده از ریتم کلیدزنی به کار گرفته است. در اینجا داده‌های مربوط به چگونگی تایپ هشت کاربر مورد استفاده قرار گرفتند. نیمی از این کاربران به عنوان کاربران مجاز و بقیه به عنوان کاربران غیر مجاز در نظر گرفته شده‌اند. با استفاده از درخت‌های تصمیم فازی، دقت کلاسیکندی (به دو کلاس کاربران مجاز و غیر مجاز) ۹۴٫۱٪ و با استفاده از درخت‌های تصمیم معمولی، دقت کلاسیکندی ۸۸٫۸٪ حاصل شده است. این درحالی است که استفاده از کلاسیکندگی‌های نزدیکترین همسایه، دقت کلاسیکندی ۸۴٫۶٪ را حاصل نموده است. علاوه بر این، به کمک درخت‌های تصمیم فازی، علاوه بر دستیابی به یک نرخ تصدیق هویت بسیار خوب، مؤثرترین خصیصه‌ها برای تصدیق هویت کاربران با استفاده از ریتم کلیدزنی نیز مورد شناسایی قرار گرفتند.

همچنین، [15] درخت‌های تصمیم فازی را به منظور تشخیص علت خرابی پتانسیومترهای (Potential Transformers) حاوی روغن به کار گرفته است. به مرور زمان که بخشی از این دستگاه‌ها دچار خرابی می‌شوند، در آنها گازهای مختلفی تولید شده و درون روغن حل می‌شوند. به مدت یک سال، برای دستگاه‌های آسیب‌دیده، روی روغن و نیز گازهای نشت شده در آن تحلیل انجام گرفته، در نتیجه، مقادیری داده به دست آمده است که به علت وجود تفاوت میان دستگاه‌ها از لحاظ اندازه، سازنده، شرایط استفاده و غیره مقادیر زیادی ناسازگاری در آنها وجود دارد. در یک پتانسیومتر، وجود گاز در روغن، نشانهٔ بروز خرابی در پتانسیومتر می‌باشد که ۱۷ دلیل شناخته شده برای وقوع آن وجود دارند. برخی از این دلایل ارتباط نزدیکتری با هم داشته، در یک تقسیم‌بندی دیگر چهار علت عمده برای خرابیها در نظر گرفته شده است. با استفاده از درخت‌های تصمیم فازی برای تشخیص علت خرابی، در حالت ۴ کلاسه دقت کلاسیکندی ۸۷٪ و در حالت ۱۷ کلاسه دقت ۶۳٫۵٪ حاصل گردیده است. از میان کاربردهای مهم دیگر درخت‌های تصمیم فازی، میتوان به مواردی همچون: پیش‌بینی درجه حرارت خروجی گرم‌کننده‌ها [22]، برنامه‌ریزی حرکت [23]، و نیز ارزیابی امنیت در سیستم‌های قدرت [24] اشاره نمود.

جمع‌بندی

در این مقاله روش یادگیری فازی با استفاده از درخت‌های تصمیم فازی به صورتی کلی معرفی گردید. ترکیب درخت‌های تصمیم نمادین با منطق فازی، قدرت آنها را در اخذ دانش نادقیق از روی نمونه‌های مغشوش، ناسازگار، دارای خطا، دارای مقادیر نامعلوم و دارای مقادیر مبهم افزایش داده، و این درحالی است که آنها با حفظ ساختار درختی خود، هنوز هم از

قابلیت درک بالایی توسط انسان برخوردار می‌باشند. به دلیل وجود چنین مزایایی کاربردهای جالبی برای آنها یافت شده است که در این مقاله به چند نمونه از آنها اشاره شد. درخت‌های تصمیم فازی با وجود برتری محسوس نسبت به درخت‌های تصمیم معمولی، هنوز هم به تحقیقات بیشتری نیازمند می‌باشند تا ضمن کاستن از اندازه درختان تولید شده، دقت آنها افزایش باز هم بیشتری پیدا نماید [20].

مراجع

- [1] Kasabov N. K., (1996), "Learning fuzzy rules and approximate reasoning in fuzzy networks and hybrid systems," Fuzzy Sets and Systems, Vol. 82, Pages 135-149.
- [2] Quinlan J.R., (1986), "Induction on decision trees," Machine Learning, Vol. 1, Pages 81-106.
- [3] Quinlan J.R., (1993), "C4.5: Programs for Machine Learning," San Mateo, CA: Morgan Kaufmann.
- [4] Janikow, C.Z., "Exemplar Learning in Fuzzy Decision Trees," Proc. 5th IEEE Int. Conf. on Fuzzy Systems, 1996, Pages 1500-1505.
- [5] Breiman L., Friedman J.H., Olshen R.A., and Stone C.J., (1984), "Classification and Regression Trees," Belmont, CA: Wadsworth.
- [6] Quinlan J.R., (1986), "The Effect of Noise on Concept Learning," Machine Learning II, Michalski R., Carbonell J. and Mitchell T. (eds), San Mateo, CA: Morgan Kaufmann.
- [7] Quinlan J.R., "Unknown Attribute-Values in Induction," Proc. 6th Int. Workshop on Machine Learning, 1989, Pages 164-168.
- [8] Zadeh L.A., (1975), "Fuzzy Logic and Approximate Reasoning," Synthese 30, Pages 407-428.
- [9] Drinakov D., Hellendoorn H. and Reinfrank M., (1993), "An Introduction to Fuzzy Control," Springer-Verlag.
- [10] Chang R.L.P, Pavlidis T., (1977), "Fuzzy Decision Tree Algorithms," IEEE Trans. on Systems, Man and Cybernetics, Vol. 7, No. 1, Pages 28-35.
- [11] Weber R., "Fuzzy-ID3: a class of methods for automatic knowledge acquisition", Proc. 2nd Int. Conf. on Fuzzy Logic and Neural Networks, Lizuka, Japan, 1992, Pages 265-268.
- [12] Maher P.E., Clair D. St., "Uncertain reasoning in an ID3 machine learning framework," Proc. 2nd IEEE Int. Conf. on Fuzzy Systems, San Francisco, CA, 1993, Pages 7-12.
- [13] Jang J., "Structure determination in fuzzy modeling: A fuzzy CART approach," Proc. 3rd IEEE Int. Conf. on Fuzzy Systems, 1994, Pages 480-488.
- [14] Janikow C.Z., (1998), "Fuzzy Decision Trees: Issues and Methods", IEEE Trans. on Systems, Man and Cybernetics-Part B: Cybernetics, Vol. 28, No. 1, Pages 1-14.
- [15] Umanol M., Okamoto H., Hatono I., Tamura H., Kawachi F., Umedzu S., Kinoshita J., "Fuzzy decision trees by fuzzy ID3 algorithm and its application to diagnosis systems," Proc. 3rd IEEE Int. Conf on Fuzzy Systems, 1994, Vol.3, Pages 2113-2118.
- [16] Koen-Myung Lee, Kyung-Mi Lee, Jee-Hyong Lee ,and Hyung Lee-Kwang, "A fuzzy decision tree induction method for fuzzy data," Proc. 8th IEEE Int. Conf. on Fuzzy Systems, 1999, Vol. 1, Pages 16-21.
- [17] Yuan Y. and Shaw M. J., (1995), "Induction of fuzzy decision trees," Fuzzy Sets and Systems, Vol.69, Pages 125-139.
- [18] Suárez A. and Lutsko J.F., (1999), "Globally optimal fuzzy decision trees for classification and regression," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.21, No. 12, Pages 1297-1311.
- [19] Olaru C. and Wehenkel L., (2003), "A complete fuzzy decision tree technique," Fuzzy Sets and Systems, Vol.138, No. 2, Pages 221-254.
- [20] Tsang E. C. C., Wang X. Z., and Yeung D. S., (2000), "Improving Learning Accuracy of Fuzzy Decision Trees by Hybrid Neural Networks," IEEE Trans. on Fuzzy Systems, Vol. 8, No. 5, Pages 601-614.
- [21] Goodarzi M.R. and Kangavari M.R., "Using Fuzzy Decision Trees for Authentication via Keystroke Timings," to be appeared in 4th Int. Conf. on Intelligent Systems Design and Applications (ISDA), Aug 2004.
- [22] Tani T. and M. Sakoda, "Fuzzy modeling by ID3 algorithm and its application to prediction of outlet temperature," Proc IEEE Int. Conf on Fuzzy Systems, San Diego, CA, Mar. 1992, Pages 923-930.
- [23] Shibata T., Abe T., Tanie K., and Nose M, "Motion planning of a redundant manipulator-criteria of skilled operators by fuzzy-ID3 and GMDH and optimization by GA," Proc. IEEE Int. Conf. on Fuzzy Systems, Yokohama, Japan, Mar. 1995, Pages 99-102.
- [24] Boyen X. and Wehenkel L., (1999), "Automatic induction of fuzzy decision trees and its application to power system security assessment," Fuzzy Sets and Systems, Vol. 102, Pages 3-19, 1999.