علوم و مهندسی کامپیوتر
نشریه علمی پژوهشی انجمن کامپیوتر ایران
مجلد ۳، شماره ۳ (الف)، پائیز ۱۳۸۴
صفحات ۹۵-۱۰۷
مقاله مروری

ترکیب طبقه‌بندها: ایجاد گوناگونی و قواعد ترکیب

سیدحسن نبوی کریزی　　احسان اله کبیر

بخش مهندسی برق، دانشگاه تربیت مدرس، تهران، ایران

چکیده

استفاده از ترکیب نتایج چند طبقه‌بند یکی از روشهای افزایش کارایی سیستمهای بازشناسی الگو است که در سالهای اخیر محققین زیادی به آن پرداخته اند. برای آنکه ترکیب نتایج طبقه‌بندها مفید واقع شود باید طبقه‌بندهای پایه ضمن برخورداری از کارایی قابل قبول، با یکدیگر متفاوت بوده و قاعده ترکیب مناسبی برای تلفیق نتایج آنها بکار گرفته شود. قاعده ترکیب باید به گونه‌ای انتخاب شود که طبقه‌بندها نقاط ضعف یکدیگر را بپوشانند. تنها با شناخت قواعد مختلف ترکیب و ویژگیهای آنها است که می‌توان قاعده ترکیب مناسب را برای حل مسأله مورد نظر انتخاب کرد. در این مقاله کلیاتی در مورد قواعد مختلف ترکیب و روشهای ایجاد گوناگونی در طبقه‌بندهای یک سیستم مرکب ارایه می‌شود. برای فهم بیشتر روشهای ترکیب، سعی می‌شود هر روش با ذکر یک مثال ساده توضیح داده شود. قواعد بیشینه، کمینه، میانگین، حاصلضرب، رأی‌گیری، انتگرال فازی، دمپستر-شفر، فضای دانش رفتار و کلیشه تصمیم، روشهای ترکیبی هستند که بررسی می‌شوند.

کلمات کلیدی: طبقه‌بندی مرکب، طبقه‌بندهای پایه، قواعدترکیب، ترکیب طبقه‌بندها، گوناگونی درخطا، همجوشی تصمیم

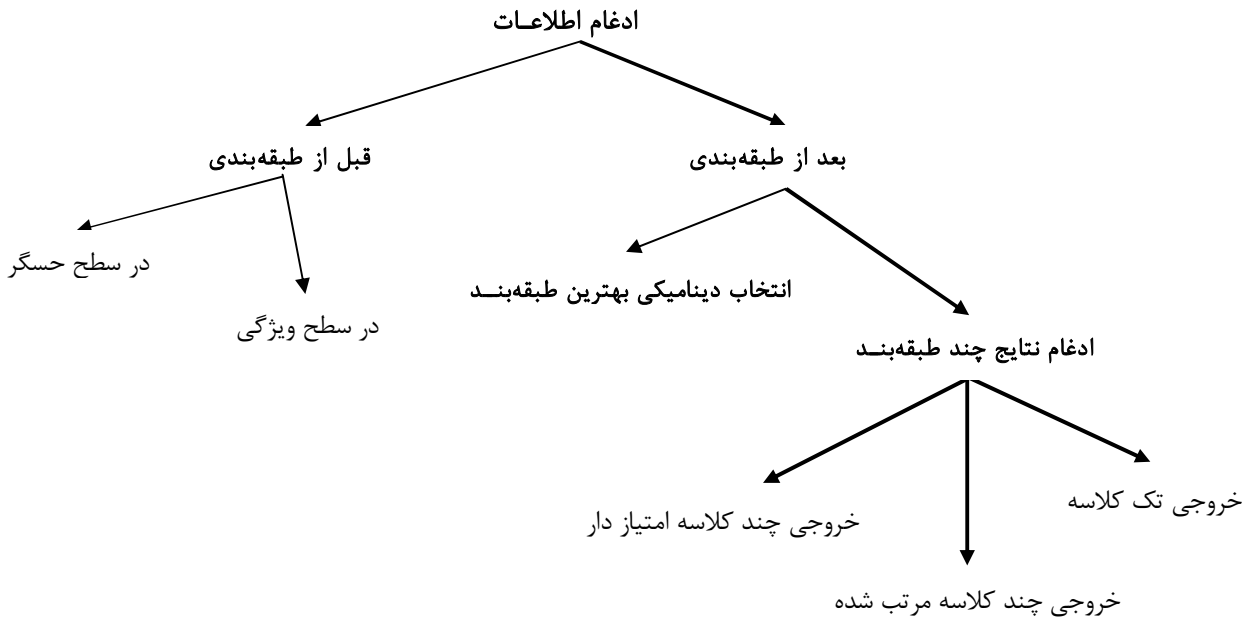
۱- مقدمه

طبقه‌بندی فرآیندی است که در آن یک الگوی ناشناخته بر اساس ویژگیهای آن، به یکی از کلاسهای شناخته شده نسبت داده می‌شود. بسیاری از الگوریتمهای یادگیری در حقیقت یک نوع جستجوی محلی انجام می‌دهند که ممکن است در کمینه‌های محلی گرفتار شوند. از آنجا که معمولاً طبقه‌بندها در یک فرایند یادگیری ساخته می‌شوند، امکان گرفتار شدن آنها در کمینه‌های محلی وجود دارد. اجرای الگوریتمهای یادگیری تحت شرایط متفاوت، یادگیری دسته جمعی[۱] ، روشی است برای آنکه بتوان تقریب بهتری از یک طبقه‌بند بهینه را فراهم کرد. در یادگیری دسته جمعی، هر الگوریتم یادگیری با توجه به مقدار پارامترهایش، به پاسخ متفاوتی برای مسأله می رسد و انتظار می‌رود با ترکیب این پاسخها، دقت طبقه‌بندی افزایش پیدا کند[۱]. به هر یک از طبقه‌بندهای یک سیستم مرکب، طبقه‌بند پایه[۲] گفته می‌شود.

استفاده از ادغام اطلاعات به منظور افزایش کارایی سیستمهای تشخیص الگو، یک مسأله متداول در علوم مختلف است. تلفیق تصمیم‌گیری برای طبقه‌بندی تصاویر چند طیفی- چند ماهواره ای[۲] ، ادغام تصاویر و داده‌های ماهواره‌ای [۴،۳] هدایت هوشمند کشتی با استفاده از ترکیب اطلاعات سنسوری[۵]، ترکیب چند سیستم خبره برای بازشناسی حروف [۶] ، پیش بینی بار با ترکیب شبکه‌های عصبی [۷]، ترکیب اطلاعات برای جداسازی بافتهای مغزی[۸] ، تشخیص هویت با ترکیب اطلاعات بیومتریک[۹] ، تشخیص نفوذ در شبکه‌های کامپیوتری[۱۰]، شناسایی کد پستی[۱۱]، بازشناسی دستنوشته[۱۲] و شناسایی گوینده[۱۳] نمونه‌هایی از کاربردهای ادغام اطلاعات هستند.

در این مقاله کلیاتی در مورد قواعد مختلف ترکیب و روشهای ایجاد گوناگونی در طبقه‌بندهای یک سیستم مرکب ارایه می‌شود. برای فهم بیشتر روشهای ترکیب، سعی می‌شود هر روش با ذکر یک مثال ساده توضیح داده شود. قواعد بیشینه، کمینه، میانگین، حاصلضرب، رأی‌گیری، انتگرال فازی، دمپستر-شفر[۳]، فضای دانش رفتار[۴] و کلیشه تصمیم[۵] ، روشهای ترکیبی هستند که بررسی می‌شوند.

ساختار مقاله به صورت زیر است. در بخش ۲ یک تصویر کلی از سطوح ادغام اطلاعات ارایه شده و جایگاه ترکیب طبقه‌بندها در آن مشخص می‌شود. در بخش ۳ اهمیت ترکیب طبقه‌بندها و نیازهای اساسی برای مفید واقع شدن آن بیان می‌شود. بخش چهارم به ایجاد گوناگونی در طبقه‌بندهای پایه می‌پردازد. بخش پنجم روشهای متداول برای ترکیب خروجی طبقه‌بندها را بیان می‌کند و بخش ششم به جمع بندی و نتیجه گیری اختصاص دارد.



شکل۱- تصویر کلی از ادغام اطلاعات

۲- سطوح ادغام اطلاعات

ادغام اطلاعات بسته به مرحله‌ای از پردازش که در آن عمل ادغام صورت می‌گیرد شامل دو سطح کلی است. ادغام اطلاعات قبل از طبقه‌بندی و ادغام اطلاعات پس ازمرحله طبقه‌بندی. شکل۱ سطوح ادغام اطلاعات را نشان می‌دهد. ادغام اطلاعات قبل از طبقه‌بندی شامل ادغام سیگنال خروجی سنسورها (ادغام در سطح پایین) و ادغام بردارهای ویژگی (ادغام در سطح متوسط) است. ادغام اطلاعات پس از طبقه‌بندی (ادغام در سطح بالا) شامل ادغام تصمیم چند طبقه‌بند و انتخاب بهترین طبقه‌بند است[۱۴]. در ادغام سطح پایین، چند سیگنال جهت ایجاد یک سیگنال با اطلاعات کاملتر ترکیب می‌شوند. در ادغام سطح متوسط، ویژگیهای مختلف مثل لبه‌ها، خطوط و گوشه‌های تصویر با هم ترکیب شده و یک ویژگی ادغام شده، بوجود می‌آید. در ادغام سطح بالا، نتایج دریافتی از چند طبقه‌بند پایه با هم ترکیب شده و تصمیم نهایی حاصل می‌شود همچنین ممکن است تصمیم نهایی به صورت انتخاب تصمیم بهترین طبقه‌بند باشد.

هرچه ادغام اطلاعات به مراحل اولیه پردازش نزدیکتر باشد، نتیجه بهتری از آن حاصل می‌شود[۹]. در این مقاله به ادغام اطلاعات پس از طبقه‌بندی یا به عبارتی ترکیب طبقه‌بندها می‌پردازیم (پیکانهای پر رنگ در شکل ۱) .

به منظور بهره گیری از نتایج چند طبقه‌بند، سه راهکار ترکیب[۶] طبقه‌بندها، همکاری[۷] طبقه‌بندها و انتخاب بهترین طبقه‌بند پیشنهاد شده است. برخی مراجع هر سه راهکار را با عنوان ترکیب طبقه‌بندها مطرح کرده‌اند[۱۵].

در راهکار ترکیب، طبقه‌بندها در یک ساختار موازی قرار دارند. الگوی ورودی به همه طبقه‌بندها اعمال می‌شود و هر کدام از آنها یک جواب برای مسأله ارایه می‌کنند. سپس با ترکیب مناسب این جوابها پاسخ نهایی بدست می‌آید. در این حالت هدف از ترکیب، بهبود دقت حل مسأله و بالابردن کارایی سیستم است.

در حالت همکاری طبقه‌بندها، یک مسأله پیچیده به چند زیر- مسأله شکسته می‌شود و هر زیر- مسأله توسط یک طبقه‌بند حل می‌شود[۱۶]. پاسخ نهایی مسأله از این پاسخها حاصل می‌شود. همچنین ممکن است بردار ویژگی به چند زیر-بردار تقسیم شده و به چند طبقه‌بند مجزا که مثلا به صورت سلسله مراتبی عمل می‌کنند اعمال شود. در این حالت از تصمیم طبقه‌بندهای لایه‌های پائین به عنوان ورودی برای طبقه‌بندهای لایه‌های بالاتر استفاده می‌شود. در حالت همکـاری نتیجه یک طبقــه‌بند می‌توانــد در مرحلـــه یادگیـــری یـا مرحلـــه تشخیص در طبقه‌بندهای دیگر استفاده شود.

در حالت انتخاب، باید پاسخ طبقه‌بندی که بیشترین شباهت را به تصمیم‌گیری صحیح دارد به عنوان پاسخ مسأله انتخاب شود، اما چون پاسخ صحیح معلوم نیست، ممکن است چند طبقه‌بند به عنوان طبقه‌بندهای منتخب منظور شوند و ترکیب نتایج آنها پاسخ نهایی باشد [۱۷].

ترکیب طبقه‌بندها، در مراجع با عناوین متفاوتی همچون ادغام طبقه‌بندها[۸] [۱۸]، همجوشی تصمیم، اختلاط خبره‌ها[۹] [۱۹] و جمع‌بندی نظرات[۱۰] [۲۰] یاد می‌شود. تفاوت نامگذاریهای فوق به نوع وابستگی بین طبقه‌بندها، نوع خروجی آنها، استراتژی تلفیق و ساختار سیستم مرکب بر می‌گردد.

۳- ترکیب طبقه‌بندها

استفاده از نتایج چند طبقه‌بند با عنوان یادگیری دسته جمعی، یک رویکرد مؤثر در یادگیری ماشینی است که در آن به منظور بهبود دقت یادگیری، نتایج طبقه‌بندها با یکدیگر ترکیب شده و یک سیستم مرکب شکل می‌گیرد. ترکیب طبقه‌بندها شامل دو بخش است. بخش اول شامل ایجاد طبقه‌بندهای پایه مناسب، انتخاب نوع طبقه‌بندها، تعداد طبقه‌بندها و ویژگی‌های مناسب برای هر طبقه‌بند است. بخش دوم شامل نحوه ترکیب خروجی طبقه‌بندها به منظور حصول بهترین نتیجه برای طبقه‌بندی الگوهاست. در ایجاد طبقه‌بندهای پایه باید توجه داشت که به نوعی گوناگون در خطا باشند تا ترکیب نتایج آنها باعث بهبود نرخ بازشناسی سیستم مرکب شود. نحوه ایجاد گوناگونی را در بخش ۴ بررسی می‌کنیم. انتخاب نوع طبقه‌بندها و تعیین تعداد آنها به نوع کاربرد و نیازمندیهای مسأله وابسته است. انتخاب ویژگی‌های مناسب برای هر طبقه‌بند باید به گونه‌ای صورت گیرد که هم نرخ بازشناسی آن طبقه‌بند به تنهایی قابل قبول باشد و هم موارد خطای آن با دیگر طبقه‌بندهای سیستم مرکب گوناگونی داشته باشد.

نحوه ترکیب خروجی طبقه‌بندها، یا به عبارت دیگر انتخاب قاعده ترکیب، یک مسأله وابسته به کاربرد است. با توجه به نیازمندیهای مسأله، قاعده ترکیب به گونه‌ای انتخاب می‌شود که نرخ بازشناسی سیستم طبقه‌بندی مرکب بیشینه باشد. قواعد مختلفی برای ترکیب نتایج خروجی طبقه‌بندها وجود دارد که متداولترین آنها را در بخش ۵ بررسی خواهیم کرد. در ادامه بحث ابتدا اهمیت ترکیب طبقه‌بندها و سپس روشهای ایجاد گوناگونی در طبقه‌بندهای یک سیستم مرکب را بررسی می‌کنیم.

۳-۱ اهمیت ترکیب طبقه‌بندها

در مسأله تشخیص الگو، ترکیب طبقه‌بندها از اهمیت ویژه‌ای برخوردار است. سه دیدگاه در این مورد عبارتند از:

الف: هر طبقه‌بند تا حد خاصی قادر به بازشناسی صحیح الگوها است و هیچ طبقه‌بندی قادر به تشخیص صحیح همه الگوها در تمام شرایط نیست. معمولاً برای یک کاربرد خاص، یک طبقه‌بند به تنهایی نرخ بازشناسی مورد نیاز را برآورده نمی‌کند و به ترکیب چند طبقه‌بند نیاز است.

ب: ویژگیهای مختلف، بازنمایی های متفاوتی از الگو هستند که هرکدام حاوی یک نوع اطلاعات مفید در مورد آن الگو هستند. برای بازشناسی یک الگو معمولاً به استخراج ویژگیهای متفاوتی نیاز است. به عنوان مثال استفاده از اثر انگشت، عنبیه چشم و صدای شخص یک مسأله متداول در تشخیص هویت[۱۱] است[۹]. ترکیب نتایج طبقه‌بندهایی که از ویژگیهای متفاوتی استفاده می‌کنند می‌تواند کارایی سیستم تشخیص الگو را بهبود بخشد. به عنوان نمونه استفاده از سه ویژگی مکانهای مشخصه[۱۲]، گشتاور زرنیکی و ویژگی ناحیه‌ای[۱۳] باعث بهبود کارایی سیستم تشخیص ارقام دستنویس فارسی در مرجع [۲۱] شده است.

پ: استخراج برخی ویژگیها، منجر به پیدایش یک بردار ویژگی بزرگ می‌شود. تحلیل بردارهای ویژگی بزرگ توسط یک طبقه‌بند، زمان پردازش را طولانی می‌کند. این موضوع در کاربردهایی نظیر سیستمهای بیومتریک مشکل ایجاد می‌کند. ترکیب طبقه‌بندها، این امکان را فراهم می‌آورد که بردارهای ویژگی با ابعاد بزرگ به بردارهای کوچکتری تقسیم شده و با طبقه‌بندهای کوچکتر و ساده تر به صورت همزمان پردازش شوند. طبقه‌بندی نهایی با ترکیب نتایج این طبقه‌بندها صورت می‌گیرد[۲۲].

کوتاه سخن اینکه با بهره گیری از ترکیب نتایج طبقه‌بندها می‌توان کارایی سیستم تشخیص الگو را، به خصوص در مورد الگوهای پیچیده بهبود بخشید.

۳-۲ نیازهای اساسی برای ترکیب طبقه‌بندها

کارایی یک سیستم مرکب، لزوماً از کارایی بهترین طبقه‌بند موجود در سیستم بهتر نیست[۲۳]. برای آنکه ترکیب نتایج خروجی طبقه‌بندها مفید واقع شود، باید نیازهای خاصی برآورده شوند.

اولین نیاز، بکارگیری یک چارچوب ریاضی مناسب برای قاعده ترکیب است تا به طریقی از نقاط قوت طبقه‌بندها استفاده و از نقاط ضعف آنها پرهیز شود. انتخاب نوع قاعده ترکیب نقشی اساسی در کارایی سیستم طبقه‌بندی مرکب دارد.

دومین نیاز، وجود تعدادی طبقه‌بند پایه با کارایی قابل قبول و نسبتاً مستقل از یکدیگر است. نرخ بازشناسی صحیح برای طبقه‌بندهای پایه باید بیشتر از حدس تصادفی باشد. نتایج تئوری[۲۴و۲۵] و تجربی [۲۶و۲۷] نشان می‌دهند که ترکیب چند طبقه‌بند، زمانی مفید واقع می‌شود که طبقه‌بندهای پایه، نرخ خطای کمی داشته باشند و خطاهای آنها با یکدیگر متفاوت باشند. دو طبقه‌بند، زمانی دارای خطاهای متفاوتی هستند که الگوهایی که طبقه‌بند اول به صورت نادرست طبقه‌بندی کرده است، با آنهایی که دومی اشتباه طبقه‌بندی کرده است متفاوت باشند یا به عبارت دیگر تصمیم های دو طبقه‌بند، ناهمبسته[۱۵] باشند. تفاوت در موارد خطا باعث می‌شود که طبقه‌بندهای پایه خطاهای یکدیگر را بپوشانند. در ادامه بحث به بررسی روشهای ایجاد طبقه‌بندهای پایه با خطاهای متفاوت می‌پردازیم.

۴- روشهای ایجاد گوناگونی

از آنجا که طبقه‌بندها در یک فرایند یادگیری ساخته می‌شوند، به منظور داشتن طبقه‌بندهای گوناگون در خطا، باید فرایند یادگیری آنها را متفاوت کرد. برای این کار روشهای متعددی پیشنهاد شده است. این روشها به دو دسته کلی تقسیم می‌شوند: روشهای ضمنی[۱۶] و روشهای صریح[۱۷].

روشهای ضمنی، با تغییرات ضمنی در فرآیند یادگیری طبقه‌بندهای پایه سعی در گوناگون کردن آنها دارند. در این روشها در طی یادگیری هیچ معیاری از گوناگونی بررسی نمی‌شود و لذا هیچگونه تضمینی برای متفاوت شدن طبقه‌بندها وجود ندارد بلکه فقط امیدوار هستیم که خطاهای طبقه‌بندهای ایجاد شده با یکدیگرمتفاوت باشند. متداولترین این روشها، انتخاب تصادفی همراه با جایگزینی نمونه‌ها از بین کلیه نمونه‌های آموزشی است که روش کیسه کردن[۱۸] نامیده می‌شود[۳۲].

روشهای صریح، با تحت تأثیر قرار دادن مسیر یادگیری طبقه‌بندهای پایه، آنها را با یکدیگر متفاوت می‌سازند. این روشها در فرایند یادگیری طبقه‌بندها، معیاری از گوناگونی اعمال می‌کنند و بر اساس آن مسیر یادگیری طبقه‌بندها را در فضای یادگیری، تغییر می‌دهند تا طبقه‌بندهایی گوناگون در خطا ایجاد شوند. روشهای تقویتی[۱۹] و روش‌های جریمه‌ای[۲۰] از مهمترین روشهای صریح برای ایجاد گوناگونی در طبقه‌بندهای پایه هستند[۲۸].

۴-۱ روشهای ضمنی

روشهای ضمنی به صورت غیرمستقیم در فرایند یادگیری طبقه‌بندهای پایه گوناگونی ایجاد می‌کنند. در ادامه به مهمترین روشهای ضمنی برای ایجاد گوناگونی اشاره می‌شود.

۴-۱-۱ گوناگونی در روش بازنمایی الگوی ورودی

متفاوت بودن روش بازنمایی الگوی ورودی می‌تواند روشی برای ایجاد گوناگونی در طبقه‌بندهای پایه باشد[۲۶]. بازنمایی‌های متفاوت برای یک الگو ممکن است مربوط به استخراج ویژگیهای متفاوت از الگو یا مربوط به استفاده از حسگرهای متفاوت برای اخذ اطلاعات مربوط به الگو باشد. استفاده از طبقه‌بندهای یکسان با بازنماییهای متفاوت برای بیان الگوی ورودی می‌تواند طبقه‌بندهای پایه متفاوتی برای سیستم مرکب ایجاد کند.

۴-۱-۲ گوناگونی در مجموعه یادگیری

یکی از روشهای ایجاد طبقه‌بندهای گوناگون در خطا، متفاوت نمودن مجموعه یادگیری آنها است. به این منظور، روشهای تقسیم تصادفی فضای ویژگی[۲۱] بین طبقه‌بندهای پایه[۲۹]، افزودن نویز به داده‌های ورودی[۳۰]، انجام تبدیلات غیر خطی بر روی بردارهای ویژگی[۳۱]، کیسه کردن[۳۲] و الگوریتم کیسه کردن نصف-نصف[۲۲] [۳۳] پیشنهاد شده‌اند. این روشها عملیاتی را به منظور تهیه نمونه‌های آموزشی متفاوت برای طبقه‌بندهای پایه انجام می‌دهند. با انتخاب مجموعه‌های یادگیری متفاوت انتظار می‌رود طبقه‌بندهای پایه متفاوت برای سیستم مرکب فراهم شود.

۴-۱-۳ گوناگونی در روش طبقه‌بندی

با بهره گیری از روشهای مختلف طبقه‌بندی، می‌توان طبقه‌بندهایی متفاوت ایجاد کرد. به عنوان نمونه بکارگیری الگوریتمهای متفاوت می‌تواند طبقه‌بندهایی ایجاد کند که دارای تخصص‌های[۲۳] متفاوتی در فضای ویژگی باشند. تفاوت در روش طبقه‌بندی می‌تواند به عنوان مثال با بهره گیری از الگوریتمهای یادگیری متفاوت، استفاده از یک الگوریتم یادگیری با پیچیدگیهای متفاوت و استفاده از یک الگوریتم یادگیری با پارامترهای مختلف محقق شود. استفاده از یک شبکه عصبی پرسپترون چندلایه به همراه یک شبکه عصبی RBF و یک درخت تصمیم[۳۴]، استفاده از شبکه عصبی پرسپترون با تعداد نرونها یا تعداد لایه‌های متفاوت، استفاده از شبکه عصبی پرسپترون با وزنهای اولیه، تعداد دفعات یادگیری، نرخ یادگیری یا تابع هزینه متفاوت[۳۵]، نمونه‌هایی از روشهای ایجاد تفاوت در روش طبقه‌بندی هستند.

۲-۴ روشهای صریح

روشهای صریح با دستکاری ماهرانه در یادگیری، اقدام به ایجاد طبقه‌بندهای پایه با خطاهای متفاوت می‌کنند. در این روشها معیاری از گوناگونی به صورت صریح یادگیری طبقه‌بندها را تحت تأثیر قرار می‌دهد. روشهای تقویتی و روشهای جریمه‌ای از مهمترین روشهای صریح برای ایجاد گوناگونی در طبقه‌بندهای پایه هستند.

۱-۲-۴ روشهای تقویتی

در روشهای تقویتی، طبقه‌بندها با استفاده ازمجموعه‌هایی با توزیع های کاملاً متفاوت آموزش می‌بینند. این روشها می‌توانند برای افزایش کارایی هر الگوریتم آموزشی به کار روند. روشهای تقویتی انواع متفاوت دارند. تقویت به وسیله پالایش[۲۴]، تقویت با نمونه برداری دوباره[۲۵]، تقویت با وزن دهی دوباره[۲۶] از جمله روشهای تقویتی هستند [۳۶]. به طور کلی در روشهای تقویتی، ابتدا به هرکدام از نمونه‌های مجموعه آموزشی اولیه، یک شانس انتخاب مساوی داده می‌شود و با انتخاب تصادفی همراه با جایگزینی، از بین این نمونه‌ها، نمونه‌های آموزشی اولین طبقه‌بند انتخاب می‌شوند. تعداد نمونه‌های آموزشی اولین طبقه‌بند با تعداد نمونه‌های آموزشی اولیه برابر است (گرچه ممکن است برخی نمونه‌های اولیه در مجموعه آموزشی اولین طبقه‌بند چند بار تکرار شده باشند)[۳۷]. پس از انتخاب اولین مجموعه آموزشی، اولین طبقه‌بند بر روی این نمونه‌ها آموزش می‌بیند و سپس بر روی کلیه نمونه‌های اولیه آزمایش می‌شود. به منظور انتخاب نمونه‌های آموزشی طبقه‌بند دوم، شانس انتخاب نمونه‌هایی که اولین طبقه‌بند آنها را بدرستی طبقه‌بندی کرده است کاهش می‌یابد و شانس انتخاب نمونه‌هایی که درست طبقه‌بندی نشده‌اند افزایش می‌یابد. به همین ترتیب برای انتخاب نمونه‌های آموزشی طبقه‌بند سوم، شانس انتخاب نمونه‌هایی که توسط دومین طبقه‌بند درست طبقه‌بندی شده‌اندکاهش یافته و شانس انتخاب آنهایی که نادرست طبقه‌بندی شده‌اند افزایش می‌یابد. این عمل برای طبقه‌بندهای بعدی نیز به همین صورت ادامه پیدا می‌کند تا زمانیکه همه نمونه‌ها درست طبقه‌بندی شوند یا به خطای قابل تحمل برسیم. ایده اصلی این روش، متمرکز کردن یادگیری بر روی نمونه‌هایی است که طبقه‌بندی آنها به سختی صورت می‌گیرد.

۲-۲-۴ روشهای جریمه‌ای

یکی از روشهای صریح برای ایجاد گوناگونی، استفاده از یک مؤلفه جریمه[۲۷] در تابع خطای طبقه‌بندهای پایه است. این مؤلفه به گونه‌ای انتخاب می‌شود که جریمه‌ای برای اشتراک خطای طبقه‌بندهای پایه در نظر گرفته شود. شکل کلی تابع خطای طبقه‌بند iام در روشهای جریمه‌ای به صورت زیر است.

$$e_i(x_p) = \frac{1}{2}\left(f_i(x_p) - d(x_p)\right)^2 + lR(x_p) \qquad (۱)$$

که در آن $e_i(x_p)$ و $f_i(x_p)$ به ترتیب خطا و خروجی طبقه‌بند iام برای بردار ویژگی x_p، $d(x_p)$بردار خروجی مطلوب و l ضریب وزنی مؤلفه جریمه $R(x_p)$است.

روشهای جریمه‌ای، با اضافه کردن یک مؤلفه جریمه به تابع خطای طبقه‌بندهای پایه مسیر یادگیری آنها را در فضای یادگیری متفاوت می‌سازند. مؤلفه جریمه معمولاً مبتنی بر عدم همبستگی خطای[۲۸] طبقه‌بندهای پایه است[۳۸]. در[۳۹] مؤلفه جریمه‌ای به صورت زیر پیشنهاد شده است.

$$R(x_p) = \sum_{j=1}^{i-1} c(j,i).\left(d(x_p) - f_i(x_p)\right).\left(d(x_p) - f_j(x_p)\right) \qquad (۲)$$

در این روش، شبکه‌ها به صورت متوالی آموزش می‌بینند و هر شبکه هنگام آموزش نه تنها انحراف پاسخش از مقدار واقعی را کمینه می‌کند بلکه شباهت خطاهایش را با شبکه‌های آموزش یافته قبل از خودش نیز حداقل می‌کند. در این روش حاصلضرب خطای شبکه‌های i و j به عنوان مؤلفه جریمه در نظر گرفته شده است. ضریب c مشخص می‌کند که خطاهای کدام شبکه‌ها باید ناهمبسته شوند. برای حالتی که هر شبکه بخواهد فقط با شبکه قبلش ناهمبسته شود، ضریب c به صورت زیر خواهد بود.

$$c(j,i) = \begin{cases} 1 & if\ i = j-1\ \ and\ i\ is\ \ even \\ 0 & otherwise \end{cases} \qquad (۳)$$

ایراد عمده این روش این است که چون شبکه‌ها به صورت متوالی آموزش می بینند، فرایند یادگیری سیستم مرکب طولانی است.

در[۴۰]، مؤلفه جریمه‌ای پیشنهاد شده است که با آن همه شبکه‌ها به صورت همزمان آموزش می‌بینند. این روش با عنوان یادگیری با همبستگی منفی[۲۹] در کاربردهای مختلفی نظیر رگرسیون[۳۰][۴۱]، تخمین سریهای زمانی[۴۰] و تقریب تابع چند متغیره[۴۲] به کار گرفته شده و در مقایسه با دیگر روشهای یادگیری دسته جمعی کارایی بهتری داشته است. اگر خروجی سیستم مرکب برای ورودی x_p ، $f_{ens}(x_p)$ ، به صورت میانگین پاسخ طبقه‌بندهای پایه فرض شود مؤلفه جریمه برای این روش به صورت زیر خواهد بود.

$$R(x_p) = -\left(f_i(x_p) - f_{ens}(x_p)\right)^2 \qquad (۴)$$

با این مؤلفه جریمه، هر شبکه در طی یادگیری نه تنها خطایش را نسبت به خروجی مطلوب کاهش می‌دهد، بلکه تنوعش با شبکه‌های دیگر را نیز افزایش می‌دهد و پارامترهای یادگیریش را بر اساس کمینه شدن خطا و بیشینه شدن گوناگونی تنظیم می‌کند.

۵- روشهای متداول برای ترکیب خروجی طبقه‌بندها

خروجی الگوریتمهای مختلف طبقه‌بندی در یکی از شکل های زیر بیان می‌شود[۴۳].

الف ـ خروجی تک کلاسه[۳۱]: در این شکل، طبقه‌بند فقط یک برچسب کلاسی، برای الگوی ورودی ارایه می‌کند.

ب ـ خروجی چند کلاسه مرتب شده[۳۲]: در این حالت، طبقه‌بند تعلق الگوی ورودی به کلاسهای مختلف را به صورت یک لیست مرتب ارایه می‌کند. کلاس با بیشترین تعلق در ابتدای لیست و کلاس با کمترین تعلق در انتهای آن قرار دارد.

ج ـ خروجی چند کلاسه امتیازدار[۳۳]: در این شکل، طبقه‌بند برای تعلق الگوی ورودی به هر کدام از کلاسها یک مقدار عددی تخصیص می‌دهد. این مقدار عددی بیانگر میزان امتیازی است که با توجه به معیار طبقه‌بندی، برای تعلق الگوی ورودی به آن کلاس محاسبه شده است.

واضح است که، خروجی تک کلاسه شامل کمترین اطلاعات و خروجی چندکلاسه امتیازدار حاوی بیشترین اطلاعات در مورد فرایند طبقه‌بندی هستند.

متداولترین تقسیم‌بندی برای روشهای ترکیب خروجی طبقه‌بندها، تقسیم‌بندی آنها بر حسب نوع خروجی الگوریتم طبقه‌بندی است که بر اساس آن روشهای ترکیب به سه دسته تقسیم می‌شوند. روشهایی چون رأی‌گیری و فضای دانش رفتار از روشهای مربوط به خروجی تک کلاسه هستند. روشهایی مانند انتگرال فازی، قاعده ضرب، قاعده بیشینه، قاعده میانگین و دمپستر-شفر از روشهای مربوط به خروجی چند کلاسه امتیازدار هستند. از روشهای مربوط به خروجی چند کلاسه مرتب شده می‌توان روش شمارش بوردا[۳۴] و بالاترین رتبه را نام برد.

فرض کنید D و Ω به ترتیب بیانگر مجموعه طبقه‌بندها و مجموعه کلاسها باشند

$$D=\{D_1,D_2,.....,D_L\},\Omega=\{w_1,w_2,.....,w_K\} \quad (۵)$$

هر طبقه‌بند D_i با دریافت بردار ویژگی $x\in R^n$، یک بردار خروجی به صورت

$$D_i(x)=[d_{i1}(x) \quad d_{i2}(x) \quad ..... \quad d_{iK}(x)]^T \quad (۶)$$

تولید می‌کند که در آن $d_{ij}(x)$ امتیاز تعلق الگوی x به کلاس w_j است که توسط طبقه‌بند D_i گزارش شده است. تلفیق خروجی طبقه‌بندها بدین معناست که با ترکیب بردارهای خروجی L طبقه‌بند، بردار تعلق الگوی x به کلاسهای K گانه را به صورت $D(x)=F(D_1(x),D_2(x),...,D_L(x))^T$ تعیین کنیم که در آن F قاعده تجمع[۳۵] است. اگر $D(x)$ به صورت زیر فرض شود

$$D(x)=[m_D^1(x) \quad m_D^2(x) \quad ..... \quad m_D^K(x)] \quad (۷)$$

پس از تعیین این بردار، کلاس الگوی x متناظر به درایه‌ای از بردار $D(x)$ خواهد بود که بزرگتر از بقیه است. به بیان ریاضی الگوی x به کلاس w_s تخصیص می‌یابد اگر :

$$m_D^s(x)\geq m_D^t(x) \quad \forall t=1,2,...,K \quad ,t\neq s \quad (۸)$$

خروجی طبقه‌بندهای مختلف می‌تواند به صورت زیر در یک ساختار ماتریسی به نام ماتریس پروفایل تصمیم[۳۶]، DP(x) ، مرتب شود[۴۴].

$$DP(x)=\begin{bmatrix} d_{11}(x) & \cdots & d_{1j}(x) & \cdots & d_{1K}(x) \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ d_{i1}(x) & \cdots & d_{ij}(x) & \cdots & d_{iK}(x) \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ d_{L1}(x) & \cdots & d_{Lj}(x) & \cdots & d_{LK}(x) \end{bmatrix} \quad (۹)$$

سطر iام ماتریس، خروجی طبقه‌بند D_i در مورد الگوی x و ستون jام ماتریس، بردار تعلق الگوی x به کلاس w_jاست که توسط طبقه‌بندهای D_1 تا D_L ارائه شده است. در مورد طبقه‌بند با خروجی تک کلاسه، مقدار $d_{ij}(x)$ صفر یا یک است. برای طبقه‌بند با خروجی چند کلاسه مرتب شده، $d_{ij}(x)$ یک عدد صحیح است که رتبه تعلق الگوی x به کلاس w_jرا نشان می‌دهد. برای طبقه‌بند با خروجی چند کلاسه امتیازدار، با توجه به مقدار شباهت الگو به کلاسهای مختلف، $d_{ij}(x)$ مقداری بین صفر تا یک دارد.

با دو رویکرد می‌توان خروجی سیستم طبقه‌بندی مرکب را با استفاده از ماتریس DP(x) بدست آورد. رویکرد اول شامل روشهایی است که فقط با توجه به ستون jام ماتریس، مقدار تعلق الگوی x به کلاس w_j را مشخص می‌کنند. این روشها با عنوان روشهای حساس به کلاس[۳۷] ذکر می‌شوند. روشهای میانگین، کمینه، بیشینه، حاصلضرب و انتگرال فازی نمونه‌هایی از روشهای حساس به کلاس هستند.

رویکرد دوم شامل روشهایی است که برای تعلق الگوی x به هر یک از کلاسها از تمام عناصر DP(x) استفاده می‌کنند. این روشها را روشهای بی‌توجه به کلاس[۳۸] می‌گویند. از جمله این روشها می‌توان روش کلیشه تصمیم و روش ترکیب دمپستر- شفر را نام برد. حجم محاسباتی روشهای این رویکرد بیشتر از روشهای رویکرد اول است. حال به معرفی روشهای این دو رویکرد می‌پردازیم.

۱-۵ روشهای حساس به کلاس

روشهای حساس به کلاس برای محاسبه تعلق الگوی x به کلاس w_j فقط از ستون jام ماتریس DP(x) استفاده می‌کنند و به بقیه اطلاعات موجود در ماتریس توجهی ندارند. این روشها عبارتند از:

۱-۱-۵ روشهای میانگین وزندار مرتب شده[۳۹]

برای استفاده از این روشها، ابتدا هر ستون ماتریس DP(x) به صورت نزولی مرتب می‌شود، بطوریکه اولین عنصر ستون iام ماتریس به صورت $a_1=\max_k(d_{ki}(x))$ و آخرین عنصر این ستون $a_L=\min_k(d_{ki}(x))$ باشد. سپس با انتخاب بردار وزنها به صورت رابطه (۱۰) عضویت الگوی x به کلاس w_iاز رابطه (۱۱) محاسبه می‌شود.

$$w=[w_1 \quad w_2 \quad ... \quad w_L]^T \quad (۱۰)$$

$$m_D^i(x)=\sum_{k=1}^{L} w_k.a_k \quad (۱۱)$$

الف- اگر بردار وزنها به صورت $w=[0,0,...0,1]^T$ انتخاب شود، کمترین مقدار ستون i به عنوان عضویت الگوی x به کلاس w_i، $m_D^i(x)$ ، منظور خواهد شد. در این حالت قاعده ترکیب، قاعده کمینه نامیده می‌شود.

ب- اگر بردار وزنها به صورت $w=[1,0,...0,0]^T$ انتخاب شود، بیشترین مقدار ستون i به عنوان عضویت الگوی x به کلاس w_iمنظور خواهد شد و قاعده ترکیب را قاعده بیشینه گویند.

ج- اگر بردار وزنها به صورت $w=[\frac{1}{L},\frac{1}{L},...,\frac{1}{L},\frac{1}{L}]^T$ انتخاب شود، میانگین مقادیر ستون i به عنوان عضویت الگوی x به کلاس w_iمنظور خواهد شد و قاعده ترکیب را قاعده میانگین گویند.

د- اگر بردار وزنها به صورت زیر انتخاب شود، مقدار میانه ستون i به عنوان عضویت الگوی x به کلاس w_iمنظور خواهد شد و قاعده ترکیب را قاعده میانه[۴۰] گویند.

$$w=[\underbrace{0,....0}_{\frac{L-1}{2}},1,\underbrace{0,....0}_{\frac{L-1}{2}}]^T \quad if \quad L \quad is \quad odd \quad (۱۲)$$

$$w=[\underbrace{0,....0}_{\frac{L-2}{2}},\frac{1}{2},\frac{1}{2},\underbrace{0,....0}_{\frac{L-2}{2}}]^T \quad if \quad L \quad is \quad even \quad (۱۳)$$

برای حالت دو کلاسی قواعد کمینه و بیشینه معادل یکدیگرند[۱].

در تمام حالات فوق، پس از محاسبه بردار $D(x)$ به صورت رابطه (۷) کلاس الگوی x متناظر با درایه‌ای از بردار خواهد بود که بزرگتر از بقیه است.

مثال ۱. فرض کنید ۵ طبقه‌بند، L=5، و سه کلاس، K=3، داریم. اگر ماتریس پروفایل تصمیم برای الگوی x به صورت زیر فرض شود:

$$DP(x)=\begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0 & 0 & 1 \\ 0.4 & 0.3 & 0.4 \\ 0.2 & 0.7 & 0.1 \\ 0.1 & 0.8 & 0.2 \end{bmatrix}$$

بکارگیری اپراتورهای فوق، نتایج زیر را برای عضویت الگوی x به کلاسهای w_1، w_2 و w_3 حاصل می‌کند.

$$Min \rightarrow m_D(x) = [0 \quad 0 \quad 0.1]^T$$
$$Max \rightarrow m_D(x) = [0.4 \quad 0.8 \quad 1]^T$$
$$Avg \rightarrow m_D(x) = [0.16 \quad 0.46 \quad 0.42]^T$$
$$Med \rightarrow m_D(x) = [0.1 \quad 0.5 \quad 0.4]^T$$

کلاس نهایی الگوی x، متناظر با درایه‌ای از بردار $m_D(x)$ است که بزرگترین مقدار را دارد. بنابراین اپراتورهای بیشینه و کمینه، الگوی x را به کلاس w_3 و اپراتورهای میانگین و میانه آنرا به کلاس w_2 تخصیص می‌دهند.

۵-۱-۲ مدلهای غیر خطی

این روشها شامل قواعد حاصلضرب، میانگین هندسی، حاصلضرب آماری[۴۱]، شبکه‌های عصبی، انتگرال فازی و روش فضای دانش رفتار هستند. در قاعده حاصلضرب، حاصلضرب مقادیر هر ستون و در قاعده میانگین هندسی، جذرحاصلضرب مقادیر هر ستون، به عنوان میزان عضویت الگوی x به کلاس متناظر با آن ستون منظور می‌شود.

در روش حاصلضرب آماری اگر $d_{ij}(x) = P(w_j|x)$، احتمال پسین تعلق الگوی x به کلاس w_j باشد، عضویت الگوی x به این کلاس از رابطه (۱۴) بدست می‌آید.

$$m_D^j(x) = \frac{\prod_{i=1}^{L} d_{i,j}(x)}{P(w_j)^{L-1}} \qquad j = 1,2,...,K \qquad (۱۴)$$

برای احتمال پیشین $P(w_j)$، معمولاً بر اساس نمونه‌های آموزشی، تخمین زیر به کار برده می‌شود .

$$\hat{P}(w_j) = \frac{N_j}{N} \qquad j = 1,2,...,K \qquad (۱۵)$$

که در آن N_j تعداد الگوهای متعلق به کلاس w_j و N تعداد کل نمونه‌های آموزشی است. این روش با فرض اینکه طبقه‌بندها مستقل از یکدیگر باشند روش بیز ساده[۴۲] را نتیجه می‌دهد.

ایراد اساسی که به قاعده حاصلضرب وارد است، مربوط به حالتی است که یکی از طبقه‌بندها به صورت نادرست تعلق الگوی ورودی به کلاس مربوطه را مقدار کوچکی گزارش کند. در این حالت بخاطر انجام عمل ضرب، میزان تعلق نهایی الگو به این کلاس مقدار کوچکی خواهد شد و الگو به صورت نادرست به کلاس دیگری تعلق خواهد گرفت. به عنوان مثال فرض کنید الگوی x متعلق به کلاس w_2 باشد ولی در طبقه‌بندی این الگو، ماتریس پروفایل تصمیم به صورت (۱۶) ظاهر شده باشد.

$$DP(x) = \begin{bmatrix} 0.1 & 0.86 & 0.2 \\ 0 & 0.68 & 0.23 \\ 0.1 & 0.001 & 0.27 \\ 0.2 & 0.7 & 0.26 \\ 0.1 & 0.8 & 0.2 \end{bmatrix} \qquad (۱۶)$$

در این صورت با استفاده از قاعده ضرب داریم :

$$m_D(x) = [0.0002 \quad 0.0003 \quad 0.0006]^T$$

بنابراین بکارگیری قاعده ضرب برای ترکیب نتایج طبقه‌بندها، الگوی x را به کلاس w_3 می‌دهد در حالیکه قواعد دیگر ترکیب، به صورت صحیح، کلاس w_2 را به الگوی x نسبت می‌دهند. برای رفع این مشکل کیتلر و همکارانش روش ضرب بهبود یافته را پیشنهاد کرده‌اند [۴۶]. در این روش چنانچه مقدار تعلق حاصل از هرکدام از طبقه‌بندها از یک مقدار آستانه، t ، کمتر باشد، این مقدار برابر با t منظور خواهد شد. فرم ریاضی روش ضرب بهبودیافته چنین است:

$$\hat{d}_{ij}(x) = \begin{cases} t & if \quad d_{ij}(x) \le t \\ d_{ij}(x) & if \quad d_{ij}(x) > t \end{cases} \qquad (۱۷)$$

شبکه‌های عصبی نیز می‌توانند به عنوان یک ابزار برای ترکیب غیرخطی نظرات طبقه‌بندها باشند. ایده اصلی در این حالت استفاده از یک روش یادگیری دو مرحله‌ای است. مرحله اول، شامل یادگیری طبقه‌بندهای پایه است. طبقه‌بندهای پایه ممکن است بر روی کل فضای ویژگی یا بر روی بخشی از زیر- فضای ویژگی تعلیم ببینند[۴۷]. این طبقه‌بندها، لایه اول شبکه را تشکیل می‌دهند. در لایه دوم یک نگاشت از فضای خروجی طبقه‌بندهای پایه به فضای برچسبهای کلاسی مطلوب صورت می‌گیرد. یادگیری در لایه دوم به مفهوم تعیین تابع نگاشت است.

۵-۱-۲-۱ انتگرال فازی

این روش، برای حالتی استفاده می‌شود که خروجی طبقه‌بندها از نوع چند کلاسه امتیازدار است. بکارگیری این روش برای ترکیب طبقه‌بندها نتایج خوبی به همراه داشته است[۴۸]. با استفاده از روش انتگرال فازی می‌توان توانایی نه تنها یک طبقه‌بند بلکه هر زیر مجموعه‌ای از طبقه‌بندها را اندازه‌گیری کرد و بدین ترتیب گروه مناسبی از آنها را در سیستم مرکب استفاده کرد[۴۹]. توانایی زیرمجموعه‌ای از طبقه‌بندها توسط اندازه لاندا- فازی[۴۳] ، g ، مشخص می‌شود. g، یک اندازه لاندا- فازی خوانده می‌شود اگر برای هر زیر مجموعه A و B از فضای D با شرط $A \cap B = f$ داشته باشیم :

$$g(A \cup B) = g(A) + g(B) + l\,g(A).g(B) \qquad l > -1 \qquad (۱۸)$$

برای هر ورودی x می‌توان K بردار با طول L را با توجه به ماتریس پروفایل تصمیم بدست آورد. هر بردار متناظر با یک کلاس است که شامل L مقدار می باشد. برای محاسبه میزان تعلق الگوی x به کلاس w_k، یعنی پیداکردن $m_D^k(x)$ ، به صورت زیر عمل می‌شود[۴۹].

۱- برای تک‌تک طبقه‌بندها، با توجه به اهمیت طبقه‌بند، چگالی‌های فازی[۴۴] $g^1, g^2, ..., g^L$ را در نظر بگیرید. دو معیار ارائه شده برای g^i به صورت $g^i = P_i$ و $g^i = \frac{P_i}{2}$ (i=1,2,..,L) هستند که در آن P_i احتمال طبقه‌بندی درست توسط طبقه‌بند D_i است. در این دو معیار، اهمیت طبقه‌بند متناسب با میزان دقت آن در نظرگرفته شده است.

۲- ریشه حقیقی و بزرگتر از ۱- ، $l > -1$ ، معادله $l + 1 = \prod_{i=1}^{L}(1 + l g^i)$ را بدست آورید.

۳- برای یک الگوی داده شده x , k امین ستون DP(x) را به صورت نزولی مرتب کنید. در اینصورت k امین ستون DP(x) به صورت زیر در خواهد آمد.

$$[d_{i_1,k}(x) \quad d_{i_2,k}(x) \quad .....d_{i_L,k}(x)]^T$$

که در آن $d_{i_1,k}(x) > d_{i_2,k}(x) > ... > d_{i_L,k}(x)$.

۴- چگالی طبقه‌بندها را متناظر با مقادیر $d_{i_j,k}(x)$ در مرحله ۳، مرتب کنید.

۵- مقدار $g(1)$ را برابر g^{i_1} قرار دهید و سپس اندازه فازی g(t) را برای $t = 2,3,.....,L$ از رابطه (۱۹) پیدا کنید.

$$g(t) = g^{i_t} + g(t-1) + l\,g^{i_t}.g(t-1) \qquad (۱۹)$$

که در آن g^{i_t} مقدار g^i در نظر گرفته شده برای طبقه‌بند tام است.

۶- درجه تعلق نهایی الگوی x به کلاس w_k را از مصالحه بین اندازه فازی g و شواهد بیان شده توسط ستون kام ماتریس DP(X) ، به کمک رابطه (۲۰) بدست آورید.

$$m_D^k(x) = \max_{t=1}^{L} \{\min \{d_{i_t,k}(x), g(t)\}\} \qquad (۲۰)$$

در مرجع[۵۰] برای محاسبه درجه تعلق الگوی x به کلاس w_kاز رابطه (۲۱) به صورت زیر استفاده شده است.

$$m_D^k(x) = d_{i_1,k}(x) + \sum_{j=2}^{L} [d_{i_{j-1},k}(x) - d_{i_j,k}(x)] g(j-1) \qquad (۲۱)$$

بردار اندازه فازی $[g(1) \quad g(2) \quad ..... \quad g(L)]^T$، ممکن است برای کلاسها و الگوهای مختلف، متفاوت باشد.

مشکل اساسی ترکیب کننده انتگرال فازی این است که اندازه فازی g برای طبقه‌بندها معلوم نیست. روش بیان شده در فوق، یعنی استفاده از داده‌ها برای تعیین اندازه لاندا- فازی، یکی از راهکارهای غلبه بر این مشکل است.

مثال۲: فرض کنید چگالی‌های فازی برای طبقه‌بندها به صورت

$$g^1 = \hat{P}_1 = 0.63$$
$$g^2 = \hat{P}_2 = 0.7$$
$$g^3 = \hat{P}_3 = 0.66$$

و ماتریس پروفایل تصمیم برای الگوی x به صورت زیر باشد.

$$DP(x) = \begin{bmatrix} 0.2 & 0.1 & 0.4 & 0.3 \\ 0.7 & 0 & 0.1 & 0.2 \\ 0.1 & 0.1 & 0.6 & 0.2 \end{bmatrix}$$

می‌خواهیم کلاس الگوی x را با استفاده از انتگرال فازی تعیین کنیم.

با حل معادله $l+1=\prod_{i=1}^{L=4}(1+lg^i)$ مقدار $l=-0.94977$ بدست می‌آید.

اگر ستونهای ماتریس DP(x) را به ترتیب نزولی مرتب کنیم و مقدار g(t) را برای t=2,3 بدست آوریم، خواهیم داشت:

Column2	g
0.1	0.66
0.1	0.99
0	1

Column1	g
0.7	0.7
0.2	0.91
0.1	1

Column4	g
0.3	0.63
0.2	0.9
0.2	1

Column3	g
0.6	0.66
0.4	0.9
0.1	1

با استفاده از رابطه (۲۰) داریم:

$$m_D^1(x) = 0.7 \qquad m_D^2(x) = 0.1$$
$$m_D^3(x) = 0.6 \qquad m_D^4(x) = 0.3$$

چنانچه از رابطه (۲۱) استفاده کنیم داریم:

$$m_D^1(x) = 1 \qquad m_D^2(x) = 0.174$$
$$m_D^3(x) = 0.878 \qquad m_D^4(x) = 0.263$$

در هر دو حالت چون مقدار عضویت الگوی x به کلاس w_1 بیشتر از بقیه است، روش ترکیب انتگرال فازی الگوی x را به کلاس w_1 تخصیص می‌دهد.

۵-۱-۲-۲ روش فضای دانش رفتار

بیشتر روشهای ترکیب از فرض استقلال طبقه‌بندها برای تصمیم‌گیری نهایی استفاده می‌کنند. این فرض همواره صحیح نیست. روش BKS به چنین فرضی برای تصمیم‌گیری نهایی نیاز ندارد. در این روش فرض بر این است که خروجی طبقه‌بندها از نوع تک کلاسه هستند.

روش BKS با ثبت نظرات طبقه‌بندها در مورد الگوهایی که کلاس آنها معلوم است، رفتار جمعی طبقه‌بندها را مدل کرده و بر اساس این مدل برچسب کلاسی الگوی ناشناخته x را مشخص می‌کند. به عبارت دیگر در این روش، تصمیم‌گیری نهایی برای کلاس یک الگو با استفاده از رفتاری که طبقه‌بندهای مختلف در هنگام یادگیری الگوها از خود نشان داده‌اند صورت می‌گیرد. اگر L طبقه‌بند و K کلاس داشته باشیم، با احتساب کلاس وازده[۴۵] برای الگوها، جدول BKS، $(K+1)^L$ ستون خواهد داشت. برای محاسبه جدول BKS به مجموعه داده‌های زیادی نیاز است. به عنوان نمونه اگر برای هر حالت ترکیب به ۱۰ نمونه داده نیاز باشد، برای حالت پنج طبقه بند و سه کلاس الگو، به تعداد ۱۰۲۴ نمونه آموزشی ($10\times(K+1)^L = 10\times(3+1)^5 = 10240$) نیاز خواهد بود. هر چه تعداد داده‌ها بیشتر باشد، دقت روش بیشترخواهد بود. واضح است که ترکیب با این روش زمانی موفق خواهد بود که طبقه بندها، الگوها را تحت تمام شرایط ممکن یاد گرفته باشند.

فرض کنید $X=\{x_1,x_2,...,x_n\}$ مجموعه آموزشی و $S_1,...,S_L$ برچسبهایی باشند که طبقه‌بندهای $D_1,.....,D_L$ به الگوی x_j تخصیص داده‌اند. بر اساس تمام ترکیبات ممکن برچسب‌های کلاسی، یک جدول جستجو[۴۶] ساخته می‌شود. در ساخت این جدول، x_j به سلولی وارد می‌شود که ترکیب برچسب‌های کلاسی آن به صورت $D_1(x_j),...,D_L(x_j)$ است. این جدول با استفاده از مجموعه داده‌های آموزشی X ساخته می‌شود. یک نمونه جدول BKS برای حالتیکه دو طبقه‌بند، L=2 ، و سه کلاس الگو، K=3 ، داشته باشیم در جدول ۱ نشان داده شده است.

پس از آنکه با استفاده از نمونه‌های آموزشی کلیه ترکیبهای ممکن در جدول ثبت شد، برچسب نهایی هر سلول یکی از حالتهای زیر را خواهد داشت:

- فقط یک برچسب کلاسی دارد که این برچسب، بیشترین تکرار را در بین اعضای X این سلول داشته است (ستونهای ۲، ۳، ۶، ۸، ۹ و ۱۰ جدول ۱).
- برچسبی ندارد. یعنی هیچکدام از اعضای X ، این نوع از خروجی طبقه‌بندها را بدست نداده‌اند (ستون ۵ جدول ۱).
- چندین برچسب کلاسی دارد اگر بیش از یک کلاس، تعداد عضو یکسان در این سلول داشته باشند(ستون ۷ جدول ۱).

برای نمونه مفهوم ستون ششم جدول این است که در بین الگوهایی که طبقه‌بندهای D_2,D_1 به ترتیب برچسبهای w_2,w_2 را به آنها تخصیص داده‌اند، یک الگو به کلاس w_1، ۱۶ الگو به کلاس w_2 و ۶ الگو به کلاس w_3 تعلق داشته‌اند. بنابراین برچسب نهایی این سلول کلاس w_2 شده است.

ستون هفتم جدول بیانگر آنستکه در بین الگوهایی که طبقه‌بندهای D_2,D_1 برای آنها به ترتیب برچسبهای w_3,w_2 را در نظر گرفته‌اند، چهار الگو به کلاس w_1، چهار الگو به کلاس w_2 و چهار الگو به کلاس w_3 تعلق داشته‌اند بنابراین برای این سلول برچسبهای w_1، w_2 و w_3 منظور شده است.

در مواردی که یک سلول بیش از یک برچسب دارد معمولا به صورت اتفاقی یکی از برچسبها به عنوان کلاس الگو در نظر گرفته می‌شود[۴۴].

جدول۱- یک نمونه جدول BKS برای دو طبقه‌بند و سه کلاس

D_1, D_2	1,1	1,2	1,3	2,1	2,2	2,3	3,1	3,2	3,3
تعداد هر کلاس	10/3/3	3/0/6	5/4/5	0/0/0	1/16/6	4/4/4	7/2/4	0/2/5	0/0/6
برچسب سلول	1	3	1,3	0	2	1,2,3	1	3	3

۵-۱-۳ مدلهای خطی

این روشها از یک ترکیب خطی از نظر تمام طبقه‌بندها برای تصمیم‌گیری نهایی استفاده می‌کنند. ضرایب ترکیب ممکن است یکسان یا متفاوت باشند. تعیین وزنهای ترکیب خطی بر اساس رفتار طبقه‌بند (مثلا با استفاده از ماتریس سردرگمی[۴۷] طبقه‌بند) [۵۱] ، بر اساس چگالی خروجی طبقه‌بند [۱۷] و بر اساس رگرسیون[۵۲] سه روش کلی هستند که در مرجع [۵۳] بررسی شده‌اند. برخی محققین با تعریف یک تابع معیار و اکسترمم نمودن آن، وزنهای ترکیب خطی را پیدا کرده‌اند. از جمله این معیارها می‌توان نرخ بازشناسی و قابلیت اعتماد به هر طبقه‌بند را نام برد[۵۴] . وزن تخصیص داده شده به نظر هر طبقه‌بند، ممکن است یک مقدار پیوسته یا یک مقدار باینری باشد. در مرجع [۵۵] برای حالت پیوسته با در نظر گرفتن نرخ خطای طبقه‌بندی سیستم مرکب به عنوان تابع معیار و کمینه کردن آن توسط الگوریتم بهینه سازی گروه ذرات[۴۸]، وزنهای بهینه ترکیب خطی محاسبه شده‌اند. برای حالتی که خروجی طبقه‌بندها از نوع تک کلاسه باشند، روشهای رأی‌گیری و روش ترکیب بیز از متداولترین روشهای وزندهی هستند که در ادامه به بررسی آنها می‌پردازیم.

۵-۱-۳-۱ روشهای رأی‌گیری

در این روشها، اظهار نظر هر طبقه‌بند در مورد کلاس الگوی ورودی، به عنوان یک رأی محسوب می‌شود و تصمیم‌گیری نهایی بر اساس آرای اخذ شده از طبقه‌بندهای مختلف صورت می‌گیرد. الگوی ورودی به کلاسی تعلق خواهد گرفت که بیشترین رأی را بیاورد. اگر طبقه‌بندها مستقل از یکدیگر بوده و نرخ بازشناسی صحیح آنها، مستقل از تعداد کلاسها، بیش از پنجاه درصد باشد، با افزایش تعداد طبقه‌بندها، روش رأی‌گیری باعث افزایش دقت طبقه‌بندی می‌شود[۵۶].

روشهای رأی‌گیری، انواع متفاوتی دارند. از جمله می‌توان به رأی‌گیری بدون وزن[۴۹]، رأی‌گیری مبتنی بر میزان اطمینان[۵۰] ، رأی‌گیری مبتنی بر رتبه[۵۱] و رأی‌گیری وزندار[۵۲] اشاره کرد.

در روشهای رأی‌گیری بدون وزن، آراء تمام طبقه‌بندها از وزن یکسانی برخوردارند یا به عبارتی ضریب وزنی همه طبقه‌بندها برابر یک است. معیار برنده شدن در این روشها، فقط تعداد آراء کسب شده است. توافق کامل[۵۳]، رأی اکثریت[۵۴] ، اکثریت مطلق[۵۵]، ترمیمی [۵۶] و رأی چند مرحله‌ای [۵۷] از جمله روشهای رأی گیری بدون وزن هستند[۵۷].

در روشهای رأی‌گیری مبتنی بر اطمینان، رأی دهنده میزان اعتماد خود به هرنامزد را بیان می‌کند. هر چقدر میزان اعتماد رأی دهندگان به یک نامزد بیشتر باشد، احتمال برنده شدن او بیشتر است. نامزدی که بیشترین اعتماد رأی دهندگان را حاصل کند، انتخاب می‌شود. لازمه استفاده از این روش در ترکیب طبقه‌بندها، این است که اولاً طبقه‌بندها میزان اعتماد خود به طبقه‌بندی صحیح الگو را ارائه کنند و ثانیاً مقدار اعتماد طبقه‌بندها به شکل صحیحی نرمالیزه شود.

در روشهای رأی‌گیری مبتنی بر رتبه، رأی دهندگان، فهرست انتخابهای خود را به صورت یک لیست مرتب ارایه می‌کنند بطوری که برترین نامزد در ابتدای لیست و بدترین آنها در انتهای آن قراردارد. در این حالت نیز، به نوعی اعتماد رأی دهندگان منظور می‌شود، ولی مانند روش قبل به نرمالیزه کردن نیاز نیست[۵۸]. در این روشها، معمولاً میانگین رتبه‌های ارایه شده توسط رأی دهندگان، به عنوان رتبه نهایی آن کاندیدا منظور شده و کاندیدای با رتبه نهایی بالاتر، به عنوان کاندیدای برتر انتخاب می‌شود.

در روشهای رأی‌گیری وزندار از معیارهای متفاوتی برای تعیین وزن تخصیص داده شده به هر طبقه‌بند استفاده می‌شود که متداولترین آنها استفاده از عملکرد طبقه‌بند بر روی نمونه‌های آموزشی (به عنوان مثال استفاده از ماتریس سردرگمی) است.

۵-۱-۳-۲ روش بیز ساده

روشهای رأی‌گیری بر اساس برچسب خروجی طبقه‌بندها عمل می‌کنند و سابقه قبلی طبقه‌بندها را در نظر نمی‌گیرند. اگر از تئوری بیز استفاده شود، سابقه قبلی طبقه‌بندها نیز مورد توجه قرار می‌گیرد. روش ترکیب بیز برای حالتی استفاده می‌شود که خروجی طبقه‌بندها از نوع تک کلاسه باشد. علاوه بر این فرض می‌شود طبقه‌بندها مستقل از یکدیگرند. به همین خاطر این روش را بیز ساده می‌گویند. در این روش ابتدا با توجه به داده‌های آموزشی برای هر طبقه‌بند D_j ، یک ماتریس سردرگمی، CM^j ، تشکیل می‌شود. عناصر این ماتریس به صورت $cm^j(k,s)$ هستند. مقدار درایه $cm^j(k,s)$ بیانگر تعداد الگوهایی است که برچسب واقعی آنها k است یعنی متعلق به کلاس w_k هستند، ولی طبقه‌بند D_j آنها را به کلاس w_s نسبت داده است. با جمع کردن مقادیر عناصر ستون sام ماتریس سردرگمی طبقه‌بند D_j ، می‌توان $cm^j(s)$ یعنی تعداد کل عناصری که این طبقه‌بند به کلاس w_s نسبت داده است را مشخص کرد.

$$cm^j(s)=\sum_{k=1}^{K} cm^j(k,s) \qquad (۲۲)$$

با استفاده از مقادیر $cm^j(s)$ و $cm^j(k,s)$ می‌توان ماتریس برچسب LM^j را بدست آورد. این ماتریس، یک ماتریس $K\times K$ است. درایه $lm^j(k,s)$ این ماتریس از رابطه (۲۳) محاسبه می‌شود.

$$Lm^j(k,s)=\hat{P}(w_k|D_j(x)=w_s)=\frac{Cm^j(k,s)}{Cm^j(s)} \qquad (۲۳)$$

اگر $S_1,S_2,...,S_L$ برچسبهایی باشند که طبقه‌بندهای D_1 تا D_L به الگوی x تخصیص داده‌اند با فرض استقلال طبقه‌بندها، تخمین احتمال اینکه برچسب واقعی الگوی x، w_i باشد، از رابطه (۲۴) محاسبه می‌شود.

$$m_D^i(x)=\prod_{j=1}^{L}\hat{P}(w_i|D_j(x)=S_j)=\prod_{j=1}^{L}Lm_{i,sj}^j \quad i=1,...,K \qquad (۲۴)$$

که در آن $\hat{P}(w_i|D_j(x)=s_j)$ها، با توجه به داده‌های آموزشی تخمین زده می‌شوند. برای این کار تعداد الگوهایی که توسط D_j برچسب s_j خورده‌اند ولی برچسب واقعی آنها w_i است بر تعداد کل الگوهای نسبت داده شده به کلاس w_j (توسط D_j) تقسیم می‌شوند.

مثال۳ : فرض کنید سه طبقه‌بند، L=3 ، و سه کلاس الگو، K=3 ، داشته باشیم و ماتریسهای سردرگمی سه طبقه‌بند با مجموعه آموزشی ۷۳ عنصری به صورت زیر باشند.

$$CM^1=\begin{bmatrix}20&1&2\\3&18&4\\2&1&22\end{bmatrix}\quad CM^2=\begin{bmatrix}21&1&1\\0&23&2\\3&1&21\end{bmatrix}\quad CM^3=\begin{bmatrix}18&2&3\\2&22&1\\1&0&24\end{bmatrix}$$

در اینصورت ماتریسهای برچسب LM^j به صورت زیر خواهند بود:

$$LM^1=\begin{bmatrix}\frac{20}{25}&\frac{1}{20}&\frac{2}{28}\\\frac{3}{25}&\frac{18}{20}&\frac{4}{28}\\\frac{2}{25}&\frac{1}{20}&\frac{22}{28}\end{bmatrix}\quad LM^2=\begin{bmatrix}\frac{21}{24}&\frac{1}{25}&\frac{1}{24}\\0&\frac{23}{25}&\frac{2}{24}\\\frac{3}{24}&\frac{1}{25}&\frac{21}{24}\end{bmatrix}$$

$$LM^3=\begin{bmatrix}\frac{18}{21}&\frac{2}{24}&\frac{3}{28}\\\frac{2}{21}&\frac{22}{24}&\frac{1}{28}\\\frac{1}{21}&0&\frac{24}{28}\end{bmatrix}$$

حال فرض کنید الگوی x به طبقه‌بندی‌های سه‌گانه $D=\{D_1,D_2,D_3\}$ اعمال شده و $D_1(x)=w_1$ ، $D_2(x)=w_2$ و $D_3(x)=w_1$ باشد، می‌خواهیم کلاس الگوی x را طبق قاعده بیز مشخص کنیم:

$$m_D^1(x)=\prod_{j=1}^{3}Lm_{1,s_j}^j=Lm_{11}^1\times Lm_{12}^2\times Lm_{11}^3=0.0274$$
$$m_D^2(x)=\prod_{j=1}^{3}Lm_{2,s_j}^j=Lm_{21}^1\times Lm_{22}^2\times Lm_{21}^3=0.0105$$
$$m_D^3(x)=\prod_{j=1}^{3}Lm_{3,s_j}^j=Lm_{31}^1\times Lm_{32}^2\times Lm_{31}^3=0.00015$$

چون $m_D^1(x)$ از $m_D^2(x)$ و $m_D^3(x)$ بزرگتر است، روش بیز ساده کلاس w_1 را به الگوی x تخصیص می‌دهد.

۵-۲ روشهای بی توجه به کلاس

این روشها از تمام عناصر ماتریس پروفایل برای محاسبه تعلق الگوی ورودی به هر یک از کلاسها، استفاده می‌کنند. واضح است که پیچیدگی محاسباتی این روشها، بیشتر از روشهای حساس به کلاس است. متداولترین این روشها، روش کلیشه تصمیم،DT ، و روش دمپستر- شفر،DS ، هستند.

۵-۲-۱ روش کلیشه تصمیم

کلیشه تصمیم هر کلاس، DT، بیانگر پاسخ طبقه‌بندها در مورد الگوهای مربوط به آن کلاس است[۴۴] . این ماتریس برای هر کلاس، با توجه به خروجی طبقه‌بندها برای نمونه‌های آن کلاس شکل می‌گیرد. به عنوان مثال روش شکل گیری کلیشه تصمیم برای کلاس w_1 ، ماتریس DT_1، به صورت زیر است.

نمونه‌های آموزشی $\{x_1,x_2,.....,x_n\}$ را که برچسبهای کلاسی آنها مشخص است، به طبقه‌بندهای $D_1(x),D_2(x),.....,D_L(x)$ اعمال می‌کنیم و با توجه به بردار خروجی آنها، ماتریس DP(x) را تشکیل می‌دهیم. از بین ماتریسهای DP(x) ایجاد شده، ماتریسهایی را که مربوط به الگوهای کلاس w_1 است جدا می‌کنیم. با متوسط‌گیری از درایه‌های نظیر به نظیر این ماتریسها، ماتریس DT_1 ایجاد می‌شود. بیان ریاضی تشکیل ماتریس $DT_i(x)=[dt_i(k,s)(x)]$ به صورت رابطه (۲۵) است[۵۹].

$$dt_i(k,s)(x)=\frac{\sum_{j=1}^{N}Ind(x_j,i).d_{ks}(x_j)}{\sum_{j=1}^{N}Ind(x_j,i)} \qquad (۲۵)$$

که در آن $k=1,2,...,L \quad , s=1,2,....,K$

$Ind(x_j,i)$ برای الگوهای x_j متعلق به کلاس w_i یک و در غیر اینصورت صفر است. $d_{ks}(x_j)$ امتیاز تعلق الگوی x_j به کلاس w_s است که توسط طبقه‌بند D_k گزارش شده است.

پس از آنکه ماتریس کلیشه تصمیم ساخته شد، به منظور تعیین عضویت الگوی ورودی به کلاس w_j از شباهت ماتریس پروفایل تصمیم با ماتریس کلیشه تصمیم کلاس w_j، استفاده می‌شود یعنی ابتدا با توجه به نظر طبقه‌بندها، ماتریس پروفایل تصمیم در مورد الگوی ورودی تشکیل شده و سپس بر مبنای شباهت عناصر ماتریس پروفایل تصمیم با ماتریس کلیشه تصمیم کلاسها، الگوی مورد نظر طبقه‌بندی می‌شود. به منظور تعیین شباهت از معیارهای متفاوتی استفاده شده است. یکی از این معیارها مربع فاصله اقلیدسی طبق رابطه (۲۶) است [۴۹].

$$m_j(x)=1-\frac{1}{L\times C}\sum_{i=1}^{L}\sum_{k=1}^{C}\left[DT_j(i,k)-d_{i,k}(x)\right]^2 \qquad (۲۶)$$

فاصله تفاضلی متقارن معیار شباهت دیگری است که در [iii] استفاده شده است.

مثال۴: فرض کنید سه طبقه‌بند، L=3 ، ودو کلاس الگو، K=2 ، داشته باشیم و ماتریس پروفایل تصمیم و کلیشه‌های مشخصه برای دو کلاس به صورت زیر باشند.

از رابطه (۲۶) خواهیم داشت:

$$DP(x)=\begin{bmatrix}0.23&0.77\\0.14&0.86\\0.21&0.79\end{bmatrix}$$

$$DT_1=\begin{bmatrix}0.85&0.15\\0.91&0.09\\0.88&0.12\end{bmatrix}\qquad DT_2=\begin{bmatrix}0.15&0.85\\0.18&0.82\\0.14&0.86\end{bmatrix}$$

$$m_1(x)=0.5246 \quad , \quad m_2(x)=0.9957$$

بنابراین روش کلیشه تصمیم، کلاس w_2 را به الگوی x تخصیص می‌دهد.

۵-۲-۲ روش‌ترکیب دمپستر- شفر

تئوری دمپستر-شفر یک ابزار قدرتمند برای بیان نایقینی است. دمپستر و شفر با لحاظ نمودن عدم قطعیت در اعتبار فرضیه‌ها، یک شکل عمومی از تئوری بیز را ارئه نمودند که در آن برای تعیین احتمال وقوع پیشامدها از بازه‌های احتمال و بازه‌های عدم قطعیت بر مبنای قرائن(شواهد) چندگانه(مثلا حاصل از خروجی چند طبقه‌بند) استفاده می‌شود. در این تئوری فرضیه به عنوان یک اظهار نظر اصولی در خصوص ماهیت موضوع مورد بررسی و نظریه به عنوان یک فرضیه یا ترکیبی از فرضیه‌ها تعریف می‌شود. در این روش بجای اختصاص احتمال به فرضیه‌ها(روش بیز) از یک اندازه(متریک) به عنوان ارزش قرائن برای تخصیص قرائن به نظریه‌ها

استفاده می‌شود[۶۱]. در استفاده از تئوری دمپستر- شفر برای ترکیب طبقه‌بندها، ابتدا با توجه به رفتار طبقه‌بندها شواهدی از آنها استخراج می‌شود که به عنوان دانش اولیه در پیدا کردن درجه عضویت الگو به هر یک از کلاسها مورد استفاده قرار می‌گیرد. در مرجع [۱۳] این شواهد با توجه به نرخ بازشناسی[۵۸]، نرخ جانشینی[۵۹] و نرخ واپس زدگی[۶۰] طبقه‌بندها محاسبه شده است.

در مرجع [۶۳] با توجه به رفتار هر طبقه‌بند در مورد نمونه‌های آموزشی هر کلاس، یک بردار مبنا برای آن کلاس محاسبه می‌شود. برای یک سیستم مرکب شامل L طبقه‌بند، L بردار مبنا برای هر کلاس بدست می‌آید که مجموعه این بردارها کلیشه تصمیم آن کلاس را می‌سازند. در مرحله تست با توجه به شباهت ماتریس پروفایل تصمیم با کلیشه تصمیم هر کلاس، شواهد مورد نیاز برای روش دمپستر–شفر بدست آمده و میزان باور و سرانجام درجه عضویت هر کلاس محاسبه می‌شود. کلاس با بیشترین درجه عضویت، کلاس الگوی ورودی خواهد بود. به طور خلاصه می‌توان الگوریتم روش ترکیب دمپستر–شفر را در این حالت به صورت زیر بیان کرد.

۱- فرض کنید DT_j^i بیانگر iامین سطر ماتریس کلیشه تصمیم برای کلاس w_j باشد. نزدیکی بین DT_j^i و $D_i(x)$ برای تمام کلاسها و تمام طبقه‌بندها، یعنی $f_{ji}(x)$ برای j=1,2,.., K ، i=1,2,...,L ، محاسبه می‌شود. $f_{ji}(x)$ در حقیقت بیانگر شباهت نظر طبقه‌بند D_i (در مورد الگوی x) به میانگین نظراتش در مورد الگوهای کلاس w_j است. یکی از تعاریف ارائه شده برای $f_{ji}(x)$ به صورت زیر است [۶۲].

$$f_{ji}(x)=\frac{\left(1+\left\|DT_j^i-D_i(x)\right\|^2\right)^{-1}}{\sum_{k=1}^{K}[(1+\|\,DT_k^i-D_i(x)\,\|^2]^{-1}} \qquad (۲۷)$$

که در آن $\|.\|$ بیانگر نرم ماتریس است. عبارت مخرج در رابطه (۲۷) به منظور نرمالیزه نمودن مقدار شباهت است بطوری که مجموع عناصر هر ستون ماتریس شباهت $f_{ji}(x)$ برابر یک باشد.

۲- برای تمام کلاسها و تمام طبقه‌بندها میزان باور[۶۱] را با استفاده از رابطه (۲۸) بدست آورید.

$$b_j(D_i(x))=\frac{f_{ji}(x).\Pi_{k\neq j}(1-f_{ki}(x))}{1-f_{ji}(x)\left[1-\Pi_{k\neq j}(1-f_{ki}(x))\right]} \qquad (۲۸)$$

۳- درجه عضویت الگوی x به کلاس w_j را طبق رابطه (۲۹) محاسبه کنید.

$$m_D^j(x)=k\prod_{i=1}^{L}b_j(D_i(x)) \qquad j=1,2,...,K \qquad (۲۹)$$

اندیس متناظر با بزرگترین مؤلفه بردار $m_D(x)$ کلاس الگوی x است. k در رابطه فوق ضریب مربوط به نرمالیزاسیون است[۴۹].

مثال۵: فرض کنید سه طبقه‌بند، L=3 ، و دو کلاس الگو، K=2 ، داشته باشیم و ماتریس پروفایل تصمیم و کلیشه‌های مشخصه برای دو کلاس به صورت زیر باشند.

با فرض آنکه مقدار k در رابطه (۲۹) برابر یک باشد با استفاده از رابطه (۲۷)

$$DP(x)=\begin{bmatrix}0.23 & 0.77\\ 0.14 & 0.86\\ 0.21 & 0.79\end{bmatrix}$$

$$DT_1=\begin{bmatrix}0.85 & 0.15\\ 0.91 & 0.09\\ 0.88 & 0.12\end{bmatrix} \qquad DT_2=\begin{bmatrix}0.15 & 0.85\\ 0.18 & 0.82\\ 0.14 & 0.86\end{bmatrix}$$

ماترس شباهت $f_{ji}(x)$ به صورت زیر بدست می‌آید.

$$f_{ji}=\begin{bmatrix}0.3641 & 0.3146 & 0.3473\\ 0.6359 & 0.6854 & 0.6527\end{bmatrix}$$

ماتریس باور با استفاده از رابطه (۲۸) به صورت زیر بدست می‌آید.

$$belief=\begin{bmatrix}0.1725 & 0.1262 & 0.1560\\ 0.5262 & 0.5989 & 0.5509\end{bmatrix}$$

درجه عضویت الگوی x به کلاسهای مختلف با استفاده از رابطه (۲۹) به صورت زیر است.

$$m_D^j(x)=\begin{bmatrix}0.0034\\ 0.1736\end{bmatrix}$$

چون عضویت الگوی x به کلاس w_2 بیشتر از مقدار آن برای کلاس w_1 است لذا روش ترکیب دمپستر- شفر کلاس w_2 را به الگوی x تخصیص می‌دهد.

۶- جمع بندی و نتیجه گیری

در این مقاله دو شرط اساسی برای‌آنکه ترکیب طبقه‌بندها مفید واقع شود بیان شد. شرط اول گوناگون بودن طبقه‌بندهای پایه و شرط دوم انتخاب قاعده ترکیب مناسب برای تلفیق نتایج خروجی طبقه‌بندهاست. این دو شرط در قالب روشهای ایجاد گوناگونی بین طبقه‌بندهای پایه و قاعد مختلف ترکیب بررسی شدند. روشهای ایجاد گوناگونی به دو دسته کلی روشهای صریح و روشهای ضمنی تقسیم شد و نمونه هایی از هر روش مطرح شدند. در بررسی روشهای مختلف برای ترکیب خروجی طبقه‌بندها معلوم شد که برخی روشهای ترکیب نظیر روش بیز و روش دمپستر-شفر سابقه قبلی طبقه‌بندها را به عنوان دانش اولیه در قاعده ترکیب استفاده می‌کنند. ولی برخی از آنها نظیر روشهای بیشینه، کمینه، میانگین، حاصلضرب و رأی‌گیری از سابقه قبلی طبقه‌بندها استفاده نمی‌کنند. در بیشتر مسایل کاربردی به خاطر حجم کم محاسبات و سادگی از روشهای دسته دوم استفاده شده است. در برخی دیگر از روشها نیاز است که از رفتار طبقه‌بند در مورد نمونه‌های یادگیری نیز اطلاع داشته باشیم و به نوعی این رفتار را مدل کنیم. معمولاً این روشها دارای محاسبات زیادتری نسبت به روشهای دسته اول هستند. در مواردی که حجم و پیچیدگی محاسباتی روش ترکیب مهم نباشد از این روشها استفاده می‌شود. روشهای انتگرال فازی، کلیشه تصمیم و روش دمپستر- شفر از جمله این روشها هستند. تجربه‌های ما در مورد استفاده از روشهای مختلف ترکیب بر روی چندین پایگاه داده نشان می‌دهد که در بین روشهای ترکیب، روش دمپستر- شفر بیشترین پیچیدگی محاسباتی و در عوض بالاترین راندمان را دارد. در حالت کلی، انتخاب روش ترکیب مناسب به کاربرد و نیازمندیهای مسأله وابسته است. امید است این مقاله شمع راهی باشد برای محققینی که علاقمند به مبحث ترکیب طبقه‌بندها هستند.

تشکر و قدردانی: نگارندگان مقاله از دقت و نظرات سازنده داوران محترم تشکر و قدردانی می‌نمایند.

مراجع

[1] F. Alimoglu and E. Alpaydin, "Combining Multiple Representations for Pen-based Handwritten Digit recognition," *ELEKTRIC: Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 9, no. 1, pp. 1-12, 2001.

[۲] ح. دهقانی ، **طبقه‌بندی تصاویر چند طیفی - چند ماهواره‌ای سنجش از دور براساس تلفیق تصمیم گیری** ، پایان نامه کارشناسی ارشد، دانشکده فنی مهندسی، دانشگاه تربیت مدرس، ۱۳۸۱.

[۳] م. کریمی آشتیانی، **ادغام تصاویر TM و SPOT با استفاده از تبدیل موجک** ، پایان نامه کارشناسی ارشد، دانشکده علوم انسانی، دانشگاه تربیت مدرس، ۱۳۷۸.

[۴] س. همایونی، **ارزیابی روشهای ترکیب داده‌های ماهواره‌ای با قدرت تفکیک مختلف جهت بهبود و دقت طبقه‌بندی مناطق شهری اصفهان** ، پایان نامه کارشناسی ارشد، دانشکده علوم پایه، دانشگاه تربیت مدرس، ۱۳۷۸.

[۵] م. نجفی، **ترکیب اطلاعات سنسوری در هدایت و مسیریابی هوشمند کشتی**، پایان نامه کارشناسی ارشد، دانشکده فنی، دانشگاه تهران، ۱۳۷۸.

[۶] ع. دهقانی، **بازشناسی حروف مجزای دستنویس فارسی با استفاده از مدل پنهان مارکف با چگالی پیوسته و ایده ترکیب چند سیستم خبره** ، پایان نامه کارشناسی ارشد، دانشکده تحصیلات تکمیلی، دانشگاه شیراز، ۱۳۷۹.

[۷] ا. فتاحی، **پیش‌بینی کوتاه‌مدت بار با استفاده از ترکیب شبکه‌های عصبی پرسپترون و کوهنن با تاکید بر روزهای سال** ، پایان نامه کارشناسی ارشد، دانشکده فنی، دانشگاه مازندران، ۱۳۷۹.

[۸] ک. مشرف، **استفاده از شبکه عصبی سلولی و نظریه ترکیب اطلاعات به منظور جداسازی بافتهای مغز از روی تصاویر MRI** ، پایان نامه کارشناسی ارشد، دانشگاه تهران، ۱۳۷۵ .

[9] A. Jain, K. Nandakumra and A. Ross, "Score Normalization in Multimodal Biometric Systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270-2285, 2005.

[10] G. Giacinto, F. Roli and L. Didaci, "Fusion of Multiple Classifier for Intrusion Detection in Computer Networks," *Pattern Recognition Letters*, vol. 24, pp. 1795 –1803, 2003.

[11] Y. Lu, C. L. Tan, "Combination of Multiple Classifiers Using Probabilistic Dictionary and Its Application to Postcode Recognition", *Pattern Recognition*, vol. 35, pp. 2823 – 2832, 2002.

[12] S. Gunter, *Multiple Classifier Systems in Offline Cursive Handwriting Recognition*, Ph.D thesis, University of Bern, 2004.

[13] K. Chen and L. Wang, "Methods of Combining Multiple Classifiers with Different Features and their Applications to Text-Independent Speaker Identification," *Int. Jour. of Pattern Recognition and Artificial Intelligence*, vol. 11, no. 3, pp. 417– 445, 1997.

[14] D. Ruta and B. Gobrys, "An Overview of Classifier Fusion Methods," *Computing and Information Systems*, vol. 7, pp. 1-10, 2000.

[15] V. Gunes and M. Menard, "Combination, Cooperation and Selection of Classifiers: a State of the Art," *Int. Jour. of Pattern Recognition and Artificial Intelligence*, vol. 17, no. 8, pp. 1303-1324, 2003.

[16] T.Windeatt and R. Ghaderi, "Binary Codes for Multi-Class Decision Combining," *Proc. SPIE*, vol. 4051, pp. 23–34, 2000.

[17] A. Verikas, A. Lipnickas, K. Malmqvist, M. Bacauskiene and A. Gelzinis, "Soft Combination of Neural Classifiers: A Comparative Study," *Pattern Recognition Letters*, vol. 20, pp. 429-444, 1999.

[18] I. Bloch , "Information Combination Operators for Data Fusion: a Comparative Review with Classification," *IEEE Trans. on Systems, Man and Cybernetics - Part A: Systems and Humans*, vol. 26, pp. 52-67, 1996.

[19] R. A. Jacobs, M. I. Jordan, S. J. Nowlan and G. E. Hinton , "Adaptive Mixtures of Local Experts," *Neural Computation*, vol. 3, pp. 79-87, 1991.

[20] J. A. Benediktsson and P. H. Swain, "Consensus Theoretic Classification Methods," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 22, pp. 688-704, 1992.

[۲۱] س. ح. نبوی کریزی و ا. کبیر ، « بهبود بازشناسی ارقام دستنویس فارسی بکمک ترکیب طبقه‌بندها با استفاده از شبکه عصبی- فازی» ، **مجموعه مقالات پنجمین کنفرانس سیستمهای فازی ایران**، ص ۳۳۷ - ۳۴۵، تهران، ۱۳۸۳ .

[22] R. O. Duda, P. E. Hart and D. G. Stork, "Pattern Classification," *John Wiley & Sons Inc.* , New York, 2001.

[23] K. Tumer and J. Ghosh, "Error Correlation and Error Reduction in Ensemble Classifiers," *Connection Science*, vol. 8, pp. 385-404, 1996.

[24] L. Hansen and P. Salamon, "Neural Network Ensembles," *IEEE Trans. on Pattern Anal. and Mach. Intelligence*, vol. 12, no. 10, pp. 993-1001, 1990.

[25] A. Krogh and J. Vedelsby, "Neural Network Ensembles, Cross Validation and Active Learning," *Advances in Neural Information Processing systems*, vol. 7, 1995.

[26] R. Duin and D. Tax, "Experiments with Classifier Combining Rules" *Proc. Int. Workshop on Multiple Classifier Systems*, LNCS 1857, Springer, pp. 16-29, 2000.

[27] L. I. Kuncheva, M. Skurichina and R. P. W. Duin, "An Experimental Study on Diversity for Bagging and Boosting with Linear Classifiers," *Information Fusion* vol. 3, pp. 245-258, 2002.

[28] G. Brown and J. Wyatt, "Diversity Creation Methods: A Survey and Categorization,"*Information Fusion Journal*, vol. 6, no. 1, pp. 5-20, 2005.

[29] M. Skurichina and R. P. W. Duin, "Bagging, Boosting and the Random Subspace Method for Linear Classifiers," *Pattern Anal. and Applications*, vol. 2, pp. 121-135, 2002.

[30] Y. Raviv and N. Intrator, "Bootstrapping With Noise: An Effective Regularization Technique," *Connection Science*, vol. 8, pp. 355-372, 1996.

[31] A. Sharkey, N. Sharkey and G. Chandroth, "Diverse Neural Net Solutions to a Fault Diagnosis Problem", *Neural Computing and Applications*, vol. 4, pp. 218-227, 1996.

[32] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123-140, 1996.

[33] L. Breiman, "Half & Half Bagging and Hard Boundary Points," Technical Report, Statistics Department, University of California, Berkeley, 1998.

[34] J. Kittler, S. A. Hojjatoleslami and T. Windeatt, "Weighting Factors in Multiple Expert Fusion," *Proc. of British Machine Vision Conference*, pp. 42-50, Essex University, Essex U.K., 1997.

[35] T. Windeatt and R. Ghaderi , "Dynamic Weighting Factors for Decision Combining," *IEE Int. Conf. On Data Fusion, Great Malvern, UK*, pp. 123-130, 1998.

[36] Y. Freund and R. E. Schapire, "Experiments With a New Boosting Algorithm," *13th Int. Conf. on Machine Learning*, pp. 148-156, 1996.

[37] H. Drucker, C. Cortes, L. D. Jackel, Y. LeCun, and V. Vapnik, "Boosting and Other Ensemble Methods," *Neural Computing*, vol. 6, pp.1289-1301, 1994.

[۳۸] س. ح . نبوی کریزی و ا. کبیر، « ایجاد گوناگونی در طبقه‌بندهای یک سیستم مرکب » ، هفتمین کنفرانس سیستمهای هوشمند ایران، دانشگاه خواجه نصیرالدین طوسی، آذر ۱۳۸۴.

[39] B. E. Rosen, "Ensemble Learning Using Decorrelated Neural Networks," *Connection Science*, vol. 8, no. 3-4, pp. 373-384, 1996.

[40] Y. Liu, *Negative Correlation Learning and Evolutionary Neural Network Ensembles*, Ph.D thesis, University of New South Wales, 1998.

[41] X. Yao, M. Fischer and G. Brown, "Neural Network Ensembles and Their Application to Traffic Flow Prediction in Telecommunications Networks," *Proc. Int. Joint Conf. on Neural Networks*, pp. 693-698, 2001.

[42] G. Brown and X. Yao, "On the Effectiveness of Negative Correlation Learning," *First UK Workshop on Computational Intelligence*, pp. 57–62, Edinburgh, Scotland, 2001.

[43] T. K. Ho, J. Hull and S. N. Srihari, "Combination of Decisions by Multiple Classifiers," *Structured Document Image Analysis*, pp. 188-202, 1992.

[44] L. I. Kuncheva, J. C. Bezdek, R. P. W. Duin, "Decision Templates for Multiple Classifier Fusion: An Cxperimental Comparison," *Pattern Recognition*, vol. 34, no.2, pp. 299-314, 2001.

[45] C. A. Shipp and L. I. Kuncheva, "Relationship Between Combination Methods and Measures of Diversity in Combining Classifiers," *Information Fusion*, vol.3, pp.135-148, 2002.

[46] F. M. Alkoot and J.Kittler, "Modified Product Fusion," *Pattern Recognition Letters*, vol. 23, pp. 957-965, 2002.

[47] C. Dietrich, G. Palm and F. Schwenker, "Decision Templates for the Classification of Bioacoustic Time Series," *Information Fusion*, vol. 4, pp. 101-109, 2003.

[48] A. Verikas and A. Lipnickas, "Fusing Neural Networks Through Space Partitioning and Fuzzy Integration," *Neural Processing Letters*, pp. 16-53, 2002.

[49] L. I. Kuncheva , "Combining Pattern Classifiers: Methods and Algorithms," *John Wiley & Sons. Inc*., 2004.

[50] M. Grabisch, "On Equivalence Classes of Fuzzy Connectives- the Case of Fuzzy Integrals," *IEEE Trans. on Fuzzy Systems*, vol. 3, no. 1, pp. 96-109, 1995.

[51] C. L. Liu, "Classifier Combination Based on Confidence Transformation," *Pattern Recognition*, vol. 38, pp. 11 – 28, 2005.

[52] V. Tresp, and M. Taniguchi, "Combining Estimators Using Non Constant Weighting Functions," *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, 1995.

[53] N. Ueda, "Optimal Linear Combination of Neural Networks for Improving Classification Performance," *IEEE Trans. On Pattern Anal. Mach. Intell*., vol. 22, no. 2, pp. 207–215, 2000.

[۵۴] س. ح . نبوی کریزی و ا. کبیر، « ارایه روشی جدید برای ترکیب طبقه‌بندها » ، **مجموعه مقالات ششمین کنفرانس سیستمهای هوشمند ایران**، دانشگاه شهید باهنر کرمان، ص ۳۱۵ - ۳۲۳، ۱۳۸۳ .

[55] S. H. Nabavi-kerizi, M. Abadi and E. Kabir, "Linear Combination of Neural Networks Using Particle Swarm Optimization," *Proc. of 11th Int. CSI Computer Conference*, vol. 2, pp. 329-333, 2006.

[56] L. Lam and C. Y. Suen, "Application of Majority Voting to Pattern Recognition: An Analysis of Its Behavior and Performance," *IEEE Trans. SMC*, vol. 27, pp. 553-568, 1997.

[57] M. van, L. Vuurpij and L. Schomaker, "An Overview and Comparison of Voting Methods for Pattern Recognition," *IWFHR*, vol. 8, pp. 195-200, 2002.

[58] T. K. Ho, J. Hull and N. Srihari , "Decision Combination in Multiple Classifier Systems," *IEEE*

Trans. on Pattern Anal. and Mach. Intelligence, vol. 16, no.1, pp.66–75, 1994.

[59] L. I. Kuncheva, R. K. Kounchev and R. Z. Zlatev, "Aggregation of Mmultiple Classification Decisions by Fuzzy Templates," *Third European Congress on Intelligent Technologies and Soft Computing*, EUFIT'95, pp. 1470-1474, 1995.

[60] L. I. Kuncheva , "Fuzzy vs. Non-fuzzy in Combining Classifiers Designed by Boosting," *IEEE Trans. on Fuzzy Systems*, vol. 11, pp. 729-741, 2003.

[۶۱] ع. ج. رشیدی، **ادغام تصمیم در سیستمهای سنجش از دور چندگانه** ، رساله دکترا، دانشکده فنی مهندسی، دانشگاه تربیت مدرس، ۱۳۸۳.

[62] G. Rogova, "Combining the Results of Several Neural Network Classifiers," *Neural Networks*, vol. 7, pp. 777-781, 1994.

---

[1] Ensemble(Committee) learning
[2] Base Classifiers
[3] Dempster-Shafer Combination
[4] Behavior-Knowledge Space (BKS)
[5] Decision Template
[6] Combination
[7] Cooperation
[8] Classifier Fusion
[9] Mixture of experts
[10] Consensus aggregation
[11] Person identification
[12] Characteristic Loci
[13] Zoning feature
[14] Framework
[15] Uncorrelated
[16] Implicit methods
[17] Explicit methods
[18] Bagging method
[19] Boosting methods
[20] Penalty methods
[21] Random subspace method
[22] Half & Half
[23] Different expertise
[24] Boosting by Filtering
[25] Boosting by Resampling
[26] Boosting by Reweighting
[27] Penalty term
[28] Error Correlation
[29] Negative Correlation Learning
[30] Regression Problems
[31] Abstract level
[32] Rank level
[33] Measurement level
[34] Borda Count Method
[35] Aggregation rule
[36] Decision Profile
[37] Class-Conscious
[38] Class- Indifferent
[39] Ordered Weighted Averaging (OWA)
[40] Median Rule
[41] Probabilistic Product
[42] Naive Bayes
[43] l - fuzzy measure
[44] Fuzzy densities
[45] Reject
[46] Look-up table
[47] Confusion Matrix
[48] Particle Swarm Optimization (PSO)
[49] Unweighted voting
[50] Confidence voting
[51] Ranked voting
[52] weighted voting
[53] Unanimity
[54] Plurality vote
[55] Majority vote
[56] Amendment vote
[57] Runoff vote
[58] Recognition Rate
[59] Substitution Rate
[60] Rejection Rate
[61] Belief degree

سید حسن نبوی کریزی در مهرماه ۱۳۶۸ پس از اتمام تحصیلات متوسطه وارد دانشکده هندسی دانشگاه فردوسی مشهد شد و در آذرسال ۱۳۷۳ با اخذ مدرک کارشناسی در مهندسی الکترونیک از این دانشگاه فارغ التحصیل شد. وی در مهرماه سال ۱۳۷۴دوره کارشناسی ارشد خود را در دانشکده مهندسی دانشگاه تربیت مدرس آغاز کرد و درشهریور ۱۳۷۷ این دوره را به اتمام رسانید. او مدت ۴ سال به عنوان مدرس آموزشکده فنی مشهد مشغول به تدریس بود. وی از مهرماه سال ۱۳۸۱ دوره دکتری مهندسی الکترونیک را در دانشگاه تربیت مدرس آغاز کرد و هم اکنون در حال گذراندن این دوره است. زمینه‌های تحقیقاتی مورد علاقه وی عبارتند از: ترکیب طبقه‌بندها، شناسایی الگو، پردازش تصویر.

آدرس پست الکترونیکی نامبرده عبارت است از:

hnabavi@modares.ac.ir

احسان‌اله کبیر کارشناسی ارشد پیوسته خود را در مهندسی برق و الکترونیک از دانشکده فنی دانشگاه تهران و دکترای خود را در مهندسی سیستمهای الکترونیک از دانشگاه اسکس در انگلستان، به رتیب در سالهای ۱۳۶۴ و ۱۳۶۹ دریافت کرد.

او اکنون دانشیار بخش مهندسی برق دانشگاه تربیت مدرس است. زمینه‌های پژوهشی مورد علاقه او عبارتند از: بازشناسی الگو به ویژه بازشناسی متون چاپی و دستنویس و بینایی ماشین

آدرس پست الکترونیکی نامبرده عبارت است از:

kabir@modares.ac.ir