

سیستم‌های دسته‌بند چندگانه: روش‌های طراحی و

قواعد ترکیب شورا

^۱ محمدعلی باقری، ^۲ غلامعلی منتظر و ^۳ احسان‌اله کبیر
^{۱ و ۲} گروه مهندسی فناوری اطلاعات، دانشکده فنی و مهندسی، دانشگاه تربیت مدرس
^۳ آزمایشگاه سیستم‌های الکترونیک، دانشگاه تربیت مدرس

چکیده:

یکی از روش‌های مناسب برای بهبود صحت دسته‌بندی، استفاده از چند دسته‌بند مختلف و سپس ترکیب نتایج خروجی آنهاست که اغلب تحت عنوان «سیستم‌های دسته‌بند چندگانه» یا «سیستم‌های شورایی» خوانده می‌شود. سیستم‌های دسته‌بند چندگانه به‌طور کلی شامل دو بخش اصلی «ایجاد شورای دسته‌بندها» و «قواعد ترکیب» آنهاست. پژوهش‌گران حوزه‌های مختلف از جمله بازشناسایی الگو، یادگیری ماشینی و آمار استفاده از سیستم‌های شورایی را بررسی کرده‌اند. نبوی کربزی و کبیر اولین مقاله مروری فارسی را در این حوزه ارائه و روش‌های ترکیب دسته‌بندها را معرفی کرده‌اند. با این حال، بررسی روند مقالات اخیر نشان می‌دهد پژوهش در حوزه سیستم‌های شورا بر روش‌های ایجاد شورای دسته‌بندها تمرکز کرده‌اند. از این‌رو در این مقاله تلاش شده است، ابتدا چارچوبی برای رویکردهای مختلف ایجاد شورای دسته‌بندها ارائه شود. براساس این ساختار، روش‌های مختلف هر رویکرد معرفی شده و در ادامه، روش‌های ترکیب دسته‌بندها به‌جمال بیان شده است. در پایان، با بررسی روند پژوهش‌های موجود، زمینه‌های تحقیقاتی مناسب در حوزه سیستم‌های شورایی ارائه شده است.

واژگان کلیدی: سیستم‌های دسته‌بند چندگانه، ترکیب دسته‌بندها، طراحی شورا، قواعد ترکیب شورا، گوناگونی

۱- مقدمه

سیستم‌های دسته‌بند چندگانه^۱، MCS که در منابع یادگیری ماشینی با نام‌های مختلفی مانند شورای دسته‌بندها^۲، کمیته یادگیرنده‌ها^۳، اختلاط خبرگان^۴ و نظریه اجماع^۵ نیز خوانده می‌شود، یکی از چهار مسیر پیشگام در حوزه یادگیری ماشینی را شکل داده است (Dietterich 1997). در حقیقت استفاده از سیستم‌های شورایی در حوزه یادگیری ماشینی، از طبیعت انسان در استفاده از نظرات مختلف برای تصمیم‌گیری‌های مهم الهام گرفته است. ما نظر افراد مختلف را وزن می‌دهیم و آنها را ترکیب می‌کنیم تا به نتیجه نهایی برسیم (Polikar 2006).

بهبود کارایی دسته‌بندی با رویکرد سیستم‌های شورایی موجب رشد روزافزون استفاده از این سیستم‌ها در حوزه‌های کاربردی بسیاری، از جمله پزشکی (Das and Sengur, 2010; Wang et al. 2010; Kuncheva and Rodriguez 2010; Bagherian, Beigy, and Irannia (Finlay Autumn, 2008 (In Persian) تحلیل ریسک (Shi et al. 2011; Peng et al. 2011)، دسته‌بندی متن (Pino-Mejías et al. 2010; 2011)، تشخیص چهره (Ebrahimpour et al. 2008; Ebrahimpour, Kabir, and Govindarajan and Yousefi 2011)، امنیت (Chandrasekaran ; Kim and Kang 2010)، یادگیری الکترونیکی (Kotsiantis, Patriarcheas, and Xenos 2010)، بویایی هوشمند (بینی الکترونیکی) (Bagheri and Montazer 2011)، تولید (Rokach 2008) و ... شده است.

شاید یکی از اولین پژوهش‌ها در حوزه سیستم‌های دسته‌بند چندگانه مقاله (Dasarathy and Sheela 1979) در

¹ multiple classifier systems (MCS)

² classifier ensembles

³ committees of learners

⁴ mixture of experts

⁵ consensus theory

می‌توانند کارایی متفاوتی را نتیجه دهند. در این موارد، ترکیب خروجی چندین دسته‌بند با میانگین‌گیری می‌تواند ریسک انتخاب نامناسب یک دسته‌بند ضعیف را کاهش دهد. متوسط‌گیری ممکن است کارایی بهترین دسته‌بند مجموعه را کاهش دهد؛ اما در عوض، ریسک انتخابی ضعیف را کم می‌کند. این موضوع به‌طور دقیق همان‌علتی است که ما با دیگر خبرگان هم‌فکری و مشورت می‌کنیم تا تصمیم بهتری بگیریم (Polikar 2006; Dietterich 2000; Kuncheva 2004).

ب) حجم بسیار زیاد داده: در کاربردهای خاص ممکن است حجم داده بسیار زیاد باشد؛ به‌طوری‌که یک دسته‌بند نتواند به‌خوبی آنها را تحلیل کند. برای نمونه در بازرسی لوله‌های انتقال گاز با استفاده از روش‌های نشت شار مغناطیسی^۴، در هر ۱۰۰ کیلومتر ۱۰ گیگابایت داده تولید می‌شود (Polikar 2006). آموزش یک دسته‌بند با این حجم بسیار زیاد داده به‌طور معمولی عملی نیست. ثابت شده است که تقسیم داده به زیرمجموعه‌های کوچک‌تر، آموزش دسته‌بندهای مختلف با افزایش‌های مختلف داده و ترکیب خروجی آنها، اغلب رویکردی مؤثرتر است (Polikar 2006).

ج) حجم کم داده: سیستم‌های دسته‌بند چندگانه همچنین می‌توانند در مسائل با تعداد بسیار کم داده نیز به‌کار گرفته شوند. دردسترس بودن مجموعه‌های کافی و مناسب از داده آموزش، بیشترین اهمیت را برای الگوریتم‌های دسته‌بندی دارد که رابطه پنهان (مکتوم) داده را یاد بگیرند. در غیاب نمونه‌های آموزش کافی، روش‌های «نمونه‌برداری دوباره»^۵ می‌تواند برای تولید زیرمجموعه‌هایی تصادفی از نمونه‌های موجود (که هم‌پوشانی نیز دارند) به‌کار گرفته شود؛ هر زیرمجموعه می‌تواند برای آموزش دسته‌بندی متفاوت استفاده شود. ثابت شده که چنین رویکردی بسیار کارا است (Polikar 2006).

د) ایجاد تقسیم و تسخیر^۶: بدون در نظر گرفتن حجم داده موجود، حل برخی مسائل خاص برای یک دسته‌بند بسیار مشکل است. به‌ویژه، مرز تصمیم^۷ که نمونه‌های دسته‌های مختلف را از هم جدا می‌کند، ممکن است

سال ۱۹۷۹ میلادی باشد که بر تقسیم فضای ویژگی با استفاده از دو یا چند دسته‌بند بحث می‌کند. با این حال، پیشرفت اصلی در این حوزه در دهه ۱۹۹۰ میلادی رخ داد. در سال ۱۹۹۰ میلادی، «هانسن»^۱ و «سالامن»^۲ نشان دادند که می‌توان عملکرد شبکه‌های عصبی را با استفاده از گروهی از شبکه‌های عصبی با پیکربندی مشابه بهبود داد (Hansen and Salamon 1990). در همان سال، پژوهش (Schapire 1990) ثابت کرد که دسته‌بندی قوی می‌تواند با ترکیب دسته‌بندهای «ضعیف» (دسته‌بندهایی ساده که صحت دسته‌بندی آنها تنها کمی بهتر از دسته‌بندی تصادفی است) به‌دست آید؛ این موضوع پایهٔ ارائهٔ الگوریتم موفق AdaBoost در سال ۱۹۹۷ میلادی شد.

نبوی کریزی و کبیر اولین مقالهٔ مروری فارسی در این حوزه را ارائه و روش‌های ترکیب دسته‌بندها را معرفی کرده‌اند (نبوی کریزی و کبیر ۱۳۸۴). با این حال بررسی روند مقالات اخیر نشان می‌دهد پژوهش در حوزه سیستم‌های شورا بر روش‌های ایجاد شورای دسته‌بندها تمرکز کرده‌اند. از این‌رو در این مقاله ابتدا چارچوبی برای رویکردهای مختلف ایجاد شورا دسته‌بندها ارائه شده است. براساس این ساختار، روش‌های مختلف هر رویکرد معرفی شده و در ادامه، روش‌های ترکیب دسته‌بندها به‌اجمال بیان شده است. در پایان، با بررسی روند پژوهش‌های موجود، زمینه‌های پژوهشی مناسب در این حوزه فعال از یادگیری ماشین ارائه شده است. شایان ذکر است تمرکز اصلی این مقاله بر دسته‌بندی با رویکرد شورا است؛ با این حال سیستم‌های شورایی می‌توانند برای مسائل خوشه‌بندی (Hong et al. 2008) و رگرسیون (Gey and Poggi 2006; Tutz and Binder 2007) نیز به‌کار گرفته شوند.

۱-۱- دلایل استفاده از سیستم‌های دسته‌بند چندگانه

الف) دلایل آماری: کارایی خوب شبکه‌های عصبی یا دیگر دسته‌بندهای خودکار^۳ برای نمونه‌های آموزش، قابلیت تعمیم آن را برای نمونه‌های آزمون تضمین نمی‌کند و این از مشکلات مهم این دسته‌بندها است. بدین معنا که مجموعه‌ای از دسته‌بندها با کارایی آموزش مشابه

⁴ magnetic flux leakage techniques

⁵ resampling

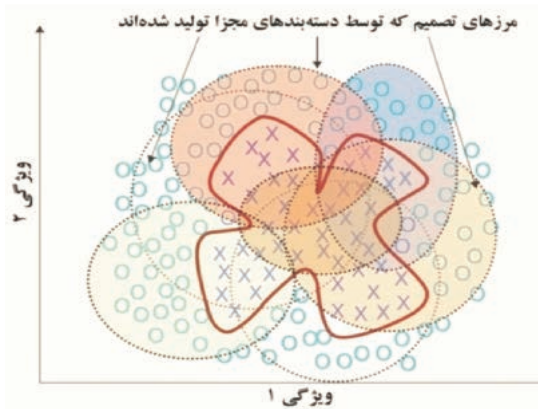
⁶ divide and conquer

⁷ decision boundary

¹ Hansen

² Salamon

³ automated classifiers



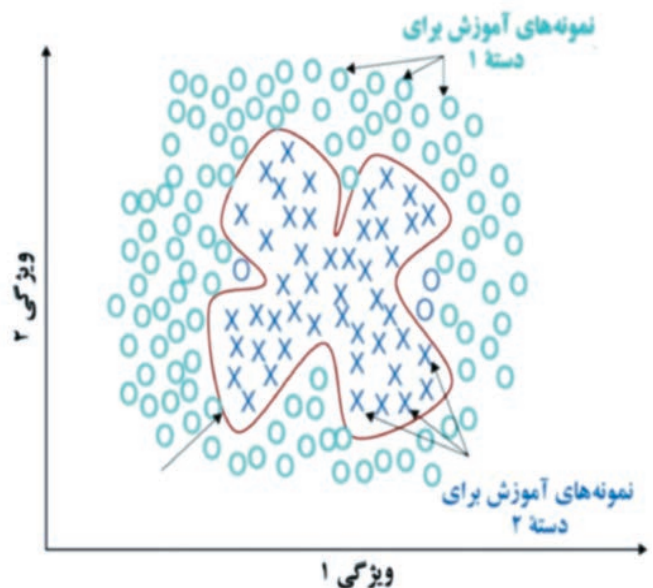
(شکل ۲): دسته‌بندهای چندگانه فضای تصمیم را پوشش می‌دهند، برگرفته از (Polikar 2006)

۲-۱- همجوشی داده: اگر مجموعه داده‌های موجود از منابع مختلفی به دست آمده باشند که طبیعت ویژگی‌هایشان متفاوت است (ویژگی‌های ناهمگن^۳)، یک دسته‌بند واحد معمولاً نمی‌تواند برای یادگیری اطلاعات موجود در همه نمونه‌ها استفاده شود. برای نمونه برای تشخیص یک بیماری عصبی، پزشک به‌طور معمول از اطلاعات مختلف مانند تصاویر MRI، اطلاعات ثبت‌نگاره مغز (الکتروآنسفالوگرامی)^۴ و آزمایش‌های خون استفاده می‌کند. هر آزمایش، نمونه‌هایی با تعداد و نوع ویژگی‌های متفاوت تولید می‌کند؛ این اطلاعات نمی‌تواند با هم برای آموزش یک دسته‌بند استفاده شود. در این‌گونه موارد، نمونه‌های از یک نوع می‌تواند برای آموزش دسته‌بندهای متفاوت استفاده شود و خروجی آنها در نهایت ادغام می‌شود. کاربردهایی را که در آن نمونه‌هایی از منابع مختلف با هم ترکیب می‌شوند تا تصمیمی آگاهانه‌تر گرفته شود کاربردهای «هم‌جوشی داده» گویند و استفاده از رویکردهای شورا برای این کاربردها نتایج خوبی را نشان داده است (Bloch 1996; Polikar 2006).

۲-۱- طراحی سیستم‌های دسته‌بند چندگانه

از نگاه کلی، طراحی سیستم‌های دسته‌بند چندگانه بر ادغام چند دسته‌بند با یک قاعده ترکیب استوار است. فرآیند طراحی شامل سه مرحله است (Yang, Qin, and Huang 2004): الف) ساخت شورا از دسته‌بندها، ب) طراحی قاعده ترکیب^۵، ج) ارزیابی کارایی. این فرآیند یادآور طراحی سیستم دسته‌بند واحد^۶، SCS، است که شامل سه مرحله است: انتخاب ویژگی‌ها،

بسیار پیچیده یا بیرون از فضای توابعی قرار گیرد که می‌تواند با مدل دسته‌بند مورد نظر ایجاد شود. به‌عنوان مثال، مسأله (شکل ۱) را با داده دویعدی و دو دسته در نظر بگیرید یک الگوریتم دسته‌بندی که تنها می‌تواند مرزهای خطی را یاد بگیرد، نمی‌تواند این مرز غیرخطی را ایجاد کند. با این حال، ترکیبی مناسب از مجموعه‌ای از این دسته‌بندهای ساده می‌تواند این مرز (یا مشابه چنین مرزهای غیرخطی) را یاد بگیرد (Polikar 2006). فرض کنید دسته‌بندی داریم که می‌تواند مرزهای بیضوی/دایره‌ای شکل را ایجاد کند. چنین دسته‌بندی نمی‌تواند مرز (شکل ۱) را یاد بگیرد. حال مجموعه‌ای از مرزهای تصمیم دایره‌ای را در نظر بگیرید که توسط شورایی از این دسته‌بندها ایجاد شده است (شکل ۲). هر دسته‌بند نمونه‌ها را با توجه به این که درون یا بیرون مرز دسته‌بند قرار می‌گیرند، نمونه‌های دسته اول (O) و نمونه‌های دسته دوم (X) برجسب می‌زند. تصمیمی براساس رأی‌گیری اکثریت^۱ از تعدادی کافی از این دسته‌بندها به سادگی می‌تواند این مرز غیرخطی پیچیده را یاد بگیرد. از سوی دیگر، این سیستم دسته‌بندی با تقسیم فضای داده به افزایش‌های کوچک‌تر که یادگیری آنها ساده‌تر است، رویکرد ایجاد تقسیم و تسخیر را دنبال می‌کند که هر دسته‌بند تنها یک افزایش ساده‌تر را یاد می‌گیرد. بدین ترتیب ترکیبی مناسب از دسته‌بندهای مختلف می‌تواند مرز تصمیم پیچیده را تقریب بزند.



(شکل ۱): مرز تصمیم پیچیده که نمی‌تواند با دسته‌بندهای خطی یا دایره‌ای یادگرفته شود، برگرفته از (Polikar 2006)

² data fusion

³ heterogeneous features

⁴ EEG (electroencephalogram)

⁵ combination rule

⁶ single classifier system

سال ۱۳۹۰ شماره ۲ پیاپی ۱۶

¹ majority voting

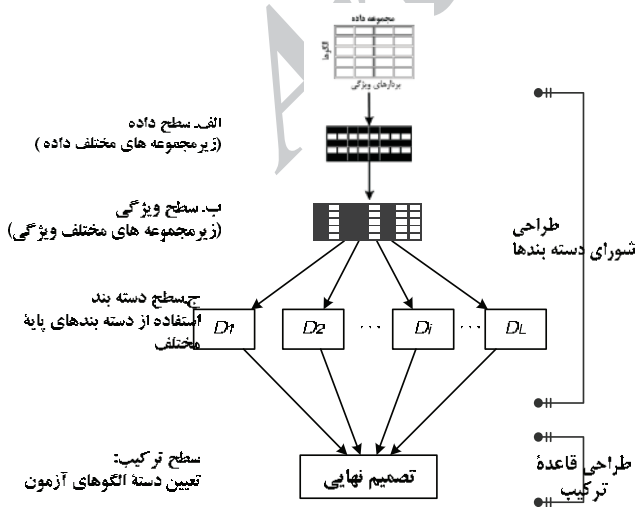
نادرست دسته‌بندی می‌کنند، متفاوت باشند. تفاوت در موارد خطای دسته‌بندی‌های پایه، باعث می‌شود دسته‌بندی‌های یکدیگر را ببوشانند و به همین علت گوناگونی در خطا، از نکات اساسی در موفقیت سیستم دسته‌بندی چندگانه است. به‌طور کلی سه روش برای ایجاد شورایی از دسته‌بندی‌های گوناگون وجود دارد:

(الف) رویکرد زیر نمونه^۵: ساخت دسته‌بندی‌های مختلف با استفاده از زیرمجموعه‌های مختلف آموزش که با نمونه‌گیری تصادفی از نمونه‌های آموزش اولیه به‌دست آمده‌اند. روش‌های نمونه‌برداری bagging (Breiman 1996) و boosting (Freund and Schapire 1996) مهم‌ترین روشها با این رویکرد هستند.

(ب) رویکرد زیر فضا^۶: ساخت دسته‌بندی‌های مختلف با زیرمجموعه‌های ویژگی مختلف.

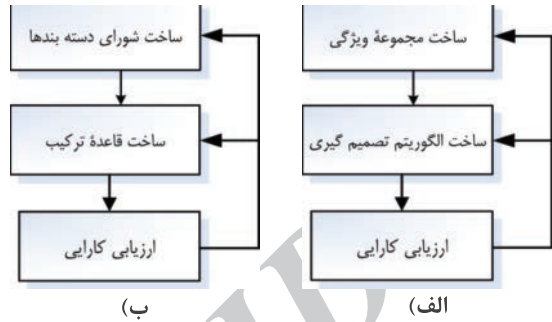
(ج) استفاده از رویکردهای یادگیری مختلف مانند شبکه عصبی، دسته‌بندی‌های بیز، درخت‌های تصمیم.

(شکل ۴) چارچوب سیستم دسته‌بندی چندگانه را نشان می‌دهد. از نگاهی دیگر، سه روش ایجاد گوناگونی، روش‌های طراحی شورای دسته‌بندی در سه سطح مختلف هستند. ساختار نگارش این مقاله بر این اساس تنظیم شده است. دو بخش اصلی سیستم‌های دسته‌بندی چندگانه (شامل طراحی شورای دسته‌بندی و ایجاد قواعد ترکیب) در بخش‌های دوم و سوم معرفی شده است. همچنین سه رویکرد ایجاد شورای دسته‌بندی، به تفکیک در زیربخش‌های بخش دوم آورده شده است. در بخش چهارم، روش‌های اندازه‌گیری گوناگونی در سیستم‌های شورایی آورده شده و در پایان، زمینه‌های پژوهشی پیشگام در این حوزه بیان شده است.



(شکل ۴): چارچوب سیستم دسته‌بندی چندگانه

ساخت الگوریتم تصمیم‌گیری و ارزیابی کارایی نهایی. فرایند طراحی این دو سیستم در (شکل ۳) نشان داده شده است (Giacinto and Roli, 2001). در این دو سیستم حلقه بازخوردی از آخرین مرحله به مراحل قبلی وجود دارد. مفهوم این حلقه در سیستم SCS روشن است. اما در MCS به این معنی است که اگر کارایی خروجی مناسب نیست، شورا یا قاعده ترکیب باید از نو ایجاد شود.



(شکل ۳): فرایند طراحی (الف) سیستم دسته‌بندی واحد و (ب) سیستم دسته‌بندی چندگانه

در SCS، مجموعه ایده‌آلی از ویژگی می‌تواند تا حد زیادی ایجاد الگوریتم تصمیم را ساده کند و الگوریتم قوی نیز می‌تواند حتی با یک مجموعه ویژگی ضعیف، نتایج خوبی را بدهد. همین موضوع در MCS نیز صادق است: مجموعه ایده‌آلی از دسته‌بندی‌ها که شامل دسته‌بندی‌های گوناگون^۱ و دقیق^۲ هستند می‌تواند تا حد زیادی ایجاد قاعده ترکیب را ساده کند و قاعده ترکیبی قوی می‌تواند حتی با شورای دسته‌بندی ضعیف، نتایج خوبی را بدهد. این دو جنبه مختلف، دو خط سیر موازی پژوهش در حوزه ترکیب دسته‌بندی‌ها هستند که در منابع یادگیری ماشینی به ترتیب تحت عنوان‌های «طراحی شورای دسته‌بندی‌ها» و «طراحی قاعده ترکیب» (یا بهینه‌سازی تصمیم^۳) آمده است. این دو موضوع اساسی در سیستم‌های شورایی در بخش‌های دوم و سوم معرفی شده است.

۱-۳- گوناگونی^۴: اساس سیستم‌های دسته‌بندی چندگانه

نتایج نظری (Hansen and Salamon Oct. 1990; Krogh and Vedelsby 1995; Hashem 1997) و تجربی (Kuncheva, Skurichina, and Duin 2000) می‌دهد زمانی یادگیری چندگانه از یادگیری بهترین دسته‌بندی پایه بهتر است که دسته‌بندی‌های پایه دارای کارایی قابل قبول بوده و گوناگون در خطا باشند. دو دسته‌بندی زمانی گوناگونی در خطا دارند که الگوهایی که به صورت

1 diverse
2 accurate
3 decision optimization
4 diversity

5 subsample
6 subspace

معیارهای مختلف گوناگونی (بخش چهارم) می‌توانند برای مقایسه شورهایی که با الگوریتم‌های متفاوت ساخته شده‌اند، به کار گرفته شوند.

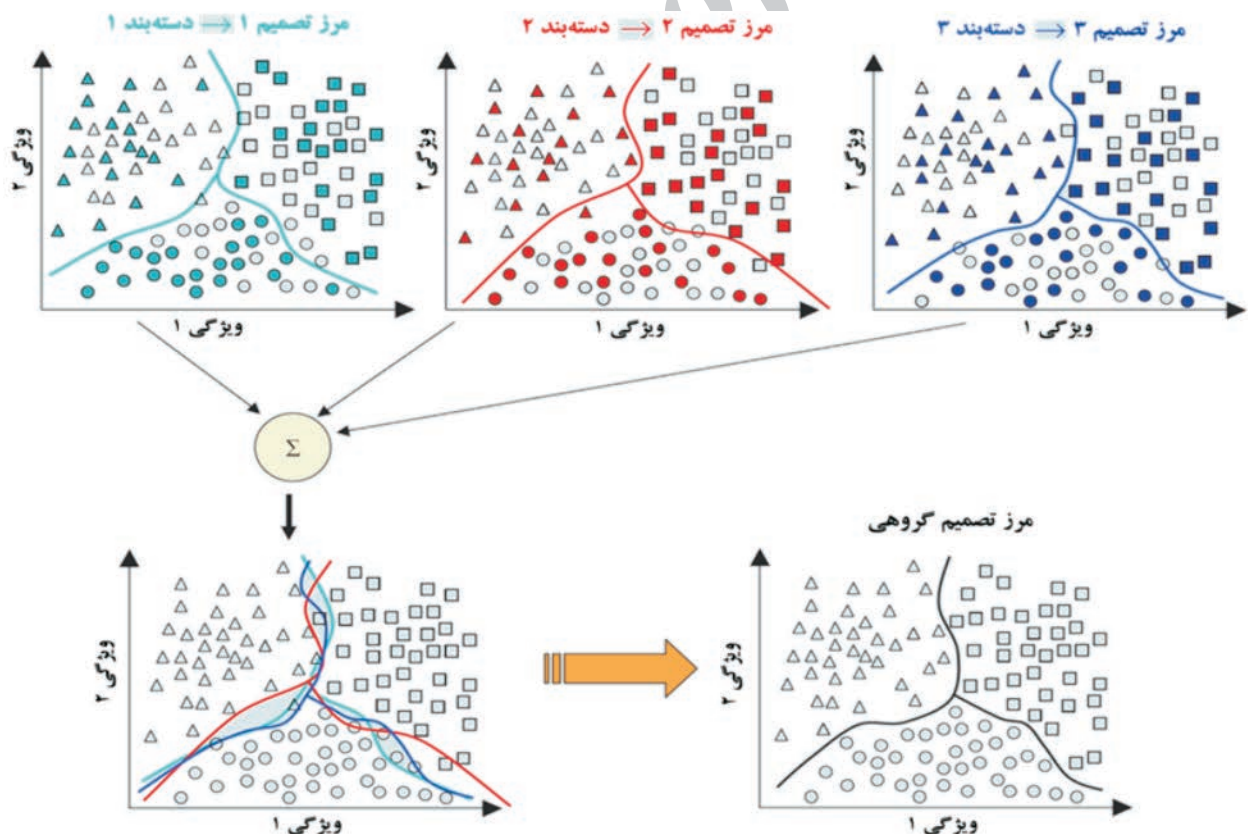
۲-۱- رویکرد زیرنمونه: سطح داده

محبوب‌ترین روش ایجاد گوناگونی، استفاده از مجموعه‌های داده مختلف برای آموزش دسته‌بندهای پایه است. این مجموعه‌ها، اغلب توسط روش‌های نمونه‌برداری دوباره حاصل می‌شود؛ به طوری که مجموعه‌های آموزش با انتخاب تصادفی، به طور معمول با جایگذاری، از میان تمام الگوهای آموزش تولید می‌شوند. این موضوع در (شکل ۵) نشان داده شده است: زیرمجموعه‌های داده تصادفی و هم‌پوشان^۳ برای آموزش سه دسته‌بند انتخاب شده است که سه مرز تصمیم متفاوت را شکل می‌دهند. آنگاه این مرزها با یکدیگر ترکیب شده‌اند تا بتوان به دسته‌بندی دقیق‌تری دست یافت.

۲- طراحی شورای دسته‌بندها

طراحی شورای دسته‌بندها، که در برخی منابع «بهینه‌سازی پوشش»^۱ نیز نامیده می‌شود، ساخت مجموعه‌ای دقیق و گوناگون از دسته‌بندهاست. بهینه‌سازی پوشش تلاش می‌کند مجموعه‌ای از دسته‌بندهای پایه را بسازد که متقابلاً مکمل همدیگر بوده و بتوانند با ترکیب همدیگر، کارایی بهتری داشته باشند. برای طراحی شورای دسته‌بندها، باید به دو سؤال وابسته به هم پاسخ داده شود: چگونه دسته‌بندهای مجزا (دسته‌بندهای پایه^۲) ایجاد شوند؟ و چگونه این دسته‌بندها با یکدیگر متفاوت شوند؟ پاسخ‌های این دو پرسش، در نهایت گوناگونی دسته‌بندها را مشخص می‌کند و در نتیجه بر کارایی کل سیستم اثر می‌گذارد (Yang, Qin, and Huang, 2004).

همچنان که اشاره شد، ایجاد شورای دسته‌بندها باید به دنبال بهبود گوناگونی باشد. با این حال الگوریتم‌های طراحی شورا، به‌طور عمومی برای بیشینه‌کردن معیار گوناگونی خاص طراحی نشده‌اند؛ بلکه گوناگونی بیشتر از طریق روش‌های نمونه‌برداری مختلف (تا حدی ابتکاری) و یا انتخاب پارامترهای مختلف آموزش به دست می‌آید (Polikar, 2006).



(شکل ۵): ترکیب دسته‌بندها با زیرمجموعه‌های داده مختلف آموزش، برگرفته از (Polikar 2006)

- 1 coverage optimization
- 2 base classifiers
- 3 overlapping

الگوریتم bagging است که چون از روش درخت تصمیم به عنوان دسته‌بندهای پایه استفاده کرده، بدین نام خوانده می‌شود (Breiman 2001). جنگل تصادفی از درخت‌های تصمیم مجزا ساخته می‌شود که نمونه‌های آموزش آن به طور تصادفی تغییر می‌کنند. این تغییر می‌تواند همانند روش bagging با نمونه‌برداری از الگوهای آموزش باشد؛ همچنین می‌توانند همانند رویکرد زیرفضای تصادفی براساس زیرمجموعه‌های ویژگی متفاوت باشد.

Boosting: این روش یکی از مهم‌ترین پیشرفت‌های سال‌های اخیر در حوزه سیستم‌های شورایی است (Freund and Schapire 1996). مشابه روش bagging، این روش مجموعه‌ای از دسته‌بندها را با نمونه‌برداری دوباره از داده ایجاد و درنهایت با «رأی اکثریت» ترکیب می‌کند. اما در روش تقویتی برای دسته‌بندهای متوالی، نمونه‌برداری دوباره با راهبرد انتخاب نمونه‌هایی است که حاوی بیشترین اطلاعات باشند. به بیان دیگر، درحالی که در روش bagging نمونه‌ها با احتمال مساوی انتخاب می‌شوند و دسته‌بندهای تولید شده مستقل هستند، روش boosting به نمونه‌هایی که اشتباه دسته‌بندی شده‌اند اهمیت بیشتری می‌دهد و دسته‌بندها را به‌طور متوالی ایجاد می‌کند. الگوریتم boosting مجموعه‌ای از وزن‌ها را برای نمونه‌های مجموعه آموزش ذخیره می‌کند. در گام اول، تمام وزن‌ها مساویند و دسته‌بند اول با این مجموعه آموزش ایجاد می‌شود. سپس وزن‌ها براساس کارایی این دسته‌بند تغییر می‌کنند؛ وزن نمونه‌هایی که اشتباه دسته‌بندی شده‌اند افزایش می‌یابد و بالعکس. لذا الگوریتم boosting، الگوریتمی تکراری است که در هر چرخه از تکرار، وزن‌ها دوباره تنظیم می‌شوند. این روند می‌تواند در مسائل یادگیری، که دارای نمونه‌هایی با پیچیدگی متفاوت هستند، کارا باشد. همچنین در روش bagging دسته‌بندها نباید پایدار باشند؛ اما در روش boosting می‌توان از دسته‌بندهای پایدار به شرط داشتن خطای دسته‌بندی کمتر از 0.5 استفاده کرد.

از مهم‌ترین رویکردهای boosting، می‌توان به رویکردهای فیلترینگ، نمونه‌برداری و وزن‌دهی مجدد^۸ اشاره کرد (Haykin and Network 2004). رویکرد اول، شامل روش‌های فیلترکردن نمونه‌های آموزش با نسخه‌های مختلف الگوریتم‌های یادگیری هستند. در رویکرد نمونه‌برداری، که روش مطرح AdaBoost نیز براساس این رویکرد ارائه شده، دسته‌بندهای مختلف با استفاده از مجموعه‌های آموزش با اندازه ثابت ایجاد می‌شوند. این نمونه‌ها در فرآیند آموزش، با

از آنجا که برای آموزش دسته‌بندها از الگوهایی به‌طور اساسی مشابه استفاده شده است، برای اطمینان از اینکه مرزهای مجزا به حد کافی متفاوت هستند، از دسته‌بندهای ناپایدار^۱ به عنوان دسته‌بندهای پایه استفاده می‌شود. زیرا این دسته‌بندهای ناپایدار می‌توانند مرزهای تصمیمی تولید کنند که حتی برای آشفتگی‌های^۲ کم در پارامترهای آموزش خود، تا حد کافی متفاوت باشند. اگر زیرمجموعه داده آموزش بدون جایگذاری ایجاد شده باشد، این روش «تقسیم k -بخشی داده»^۳ نیز نامیده می‌شود. در این روش، تمام مجموعه داده به k بلوک تقسیم می‌شود و هر دسته‌بند، یک بلوک را برای آزمون و $k-1$ بلوک دیگر را برای آموزش استفاده می‌کند. مهم‌ترین روش‌های رویکرد زیرنمونه در ادامه معرفی شده است:

Bagging: این روش توسط «بریمان»^۴ در سال ۱۹۹۶ میلادی ابداع شده (Breiman 1996) و یکی از اولین الگوریتم‌های ایجاد شورا است. همچنین این روش یکی از شهودی‌ترین و ساده‌ترین روش‌های ایجاد سیستم دسته‌بند چندگانه است که در بسیاری از مسائل از کارایی بسیار خوبی برخوردار است. روش Bagging با نمونه‌برداری با جایگذاری^۵ از مجموعه آموزش اولیه، زیرمجموعه‌های آموزش مختلفی می‌سازد که هر نمونه شانس یکسان برای انتخاب شدن دارد. اندازه هر زیرمجموعه آموزش برابر با مجموعه آموزش اولیه است. بنابراین بعضی نمونه‌ها در مجموعه‌های آموزش مختلف نیستند؛ درحالی که برخی نمونه‌ها ممکن است بیش از یک بار ظاهر شوند. این فرایند، جمع خودراه‌انداز (bootstrap aggregating) و به اختصار bagging خوانده شده است (Breiman 1996).

روش bagging برای دسته‌بندهای ناپایدار^۶ نتایج بهتری می‌دهد؛ دسته‌بندهای ناپایدار (مانند شبکه عصبی یا درخت تصمیم)، دسته‌بندهایی هستند که در صورت تغییر کوچکی در مجموعه آموزش یا پارامترهای الگوریتم، خروجی آنها تغییر زیادی می‌کند. این روش به‌ویژه برای زمانی که الگوهای موجود کم هستند، بسیار مطلوب است. تحلیل نظری و تجربی روش bagging در پژوهش (Fumera, Roli, and Serrau 2008) به تفصیل آورده شده است.

جنگل تصادفی: روش جنگل تصادفی^۷، یکی از انواع

¹ unstable classifiers

² perturbations

³ k -fold data split

⁴ Breiman

⁵ with replacement

⁶ unstable

⁷ Random Forests

⁸ reweighting

۳. خروجی مرحله آموزش: شورای دسته‌بند و وزنه‌های هر دسته‌بند β_1, \dots, β_L

آزمون: رأی‌گیری حداکثری وزن‌دار (Weighted Majority Voting)

برای نمونه برچسب نخورده x :

۴. رأی (امتیاز) هر دسته را برای این نمونه محاسبه کن:

$$\mu_t^{(x)} = \sum_{D_k(x)=w_t} \ln\left(\frac{1}{\beta_k}\right) \quad (4)$$

۵. دسته‌ای که بیشترین رأی (امتیاز) را دارد، به‌عنوان دسته نمونه x انتخاب می‌شود.

این الگوریتم توزیع وزنها، w_i^k ، برای نمونه‌های آموزش را $x_i, i = 1, \dots, N$ حفظ می‌کند. براساس این توزیع، زیرمجموعه‌های آموزش در هر گام، S_k ، برای آموزش دسته‌بند بعدی انتخاب می‌شوند. توزیع وزن نمونه‌ها در ابتدا یکنواخت است؛ لذا هر نمونه شانس انتخاب یکسانی در زیرمجموعه اول آموزش دارد. خطای آموزش دسته‌بند D_k ، ϵ_k ، نیز بر اساس توزیع وزن نمونه‌ها محاسبه می‌شود؛ به‌طوری که ϵ_k مجموع وزن نمونه‌هایی است که توسط دسته‌بند D_k اشتباه دسته‌بندی شده‌اند (رابطه ۱). سپس خطای نرمال شده دسته‌بند، β_k ، محاسبه می‌شود؛ به‌طوری که برای هر $0 < \epsilon_k < 1/2$ داریم: $0 < \beta_k < 1$.

رابطه ۳ قاعده روزآمد کردن وزن‌ها را بیان می‌کند: وزن نمونه‌هایی که به‌کمک دسته‌بند جاری صحیح دسته‌بندی شده‌اند با ضریب β_k کاهش می‌یابد؛ درحالی‌که وزن نمونه‌هایی که اشتباه دسته‌بندی شده‌اند تغییر نمی‌کند. زمانی که وزن‌های جدید، هنجار شوند (در نتیجه w^{k+1} توزیع صحیح خواهد بود)، وزن نمونه‌های اشتباه دسته‌بندی شده افزایش می‌یابد. از این‌رو در هر تکرار، الگوریتم بر روی نمونه‌هایی سخت‌تر تمرکز می‌کند. از آنجا که لازم است خطای الگوریتم یادگیری (دسته‌بند) کمتر از $1/2$ باشد، این الگوریتم تضمین می‌کند که حداقل یک نمونه که در قبل اشتباه دسته‌بندی شده بود، درست دسته‌بندی شود. زمانی که تعداد L دسته‌بند ایجاد شود، الگوریتم وارد مرحله آموزش می‌شود. برخلاف bagging و boosting، الگوریتم AdaBoost از الگوی رأی‌گیری اکثریت وزن‌دار استفاده می‌کند. ایده شهودی این موضوع آن است که دسته‌بندی‌هایی که در مرحله آموزش کارایی بهتری داشته‌اند و خطای دسته‌بندی β_k کمتری داشته‌اند، وزن بیشتری می‌گیرند. از این‌رو، معکوس مقدار β_k معیاری از کارایی است. باین حال چون β_k خطای آموزش است، اغلب نزدیک به صفر است و معکوس آن عددی بسیار بزرگ خواهد بود. برای جلوگیری از بروز مشکل ناپایداری، به‌طور معمول لگاریتم $1/\beta_k$ به‌عنوان وزن دسته‌بند استفاده می‌شود. درنهایت،

نمونه‌برداری براساس توزیع احتمال آنها انتخاب می‌شوند. در رویکرد وزن‌دهی مجتد نیز اندازه مجموعه آموزش برای دسته‌بندها ثابت است؛ اما این رویکرد، فرض می‌کند که الگوریتم یادگیری ضعیف شورا می‌تواند نمونه‌ها را با وزن‌های مختلف ببیند.

□ **AdaBoost**: در سال ۱۹۹۷ میلادی، «فروند»^۱ و «شاپیر»^۲ الگوریتم AdaBoost را معرفی کردند (Freund and Schapire, 1997) که از آن زمان توجه زیادی را در حوزه هوش مصنوعی و یادگیری ماشینی به خود جلب کرده است. در میان نسخه‌های مختلف این الگوریتم، AdaBoost.M1 و AdaBoost.R بیشترین استفاده را دارند که به ترتیب برای مسائل دسته‌بندی چنددسته‌ای^۳ و رگرسیون استفاده می‌شوند. شبه‌کد AdaBoost.M1 در ادامه آورده شده است. □

ورودی:

■ نمونه‌های آموزش $S = [(x_i, y_i)], i = 1, \dots, N$ با برچسب‌های

صحیح $y_i \in \Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$ ، ω_i

■ الگوریتم یادگیری ضعیف (WeakLearn)

■ عدد صحیح L برابر با تعداد دسته‌بندها (مساوی با تعداد تکرارها)

۱. مقدار دهی اولیه

■ تنظیم وزن نمونه‌ها $w^1 = [w_1, \dots, w_N]$ ؛ $w_j^1 \in [0, 1]$ ؛

$$\sum_{j=1}^N w_j^1 = 1$$

(معمولاً $w_j^1 = \frac{1}{N}$)

۲. انجام بده: (به‌ازای L, \dots, k)

۱. زیرمجموعه‌ای از نمونه‌های آموزش S_k براساس وزنه‌های w^k انتخاب کن.

۲. دسته‌بند D_k را با استفاده از زیرمجموعه S_k ایجاد کن.

۳. خطای وزن‌دار شورای دسته‌بند در گام k را محاسبه کن:

$$\epsilon_k = \sum_{j=1}^N w_j^k l_k^j \quad (1)$$

۱) اگر $l_k^j = 1$ اگر دسته‌بند D_k نمونه z_j را اشتباه دسته‌بندی کند و

در غیر این صورت $l_k^j = 0$

□ اگر $\epsilon_k = 0$ یا $\epsilon_k \geq 0.5$ از دسته‌بند D_k صرف‌نظر کن؛ و وزنه‌های

w_j^k را برابر $\frac{1}{N}$ قرار بده.

در غیر این صورت خطای هنجار شده شورای دسته‌بند در گام k (β_k) را محاسبه کن:

$$\beta_k = \frac{\epsilon_k}{1 - \epsilon_k} \quad (2) \quad \text{که } \epsilon_k \in (0, 0.5)$$

۴. وزن نمونه‌ها (برای گام بعد) را بروزرسانی کن:

$$w_j^{k+1} = \frac{w_j^k \beta_k^{(1-l_k^j)}}{\sum_{i=1}^N w_i^k \beta_k^{(1-l_k^i)}} \quad (3)$$

□

¹ Freund

² Schapire

³ multiclass

می‌دهند؛ در مرحله بعد، دسته‌بند D_{L+1} ، برای تخصیص وزن برای ترکیب دسته‌بندها استفاده می‌شود. دسته‌بند مرحله دوم که به‌طور معمول بر اساس شبکه عصبی است، صرفاً برای توزیع وزن به دسته‌بندهای مرحله اول است و شبکه میانجی⁴ نامیده می‌شود. نکته قابل توجه در این روش این است که ورودی این شبکه، داده‌های آموزش است و لذا وزن‌های قاعده ترکیب براساس نمونه‌های مسأله به‌دست می‌آیند و قاعده ترکیب پویا خواهد بود. آموزش شبکه میانجی به‌طور معمول با استفاده از روش بیشترین شیب⁵ یا الگوریتم بیشینه‌سازی مورد انتظار⁶ است.

روش اختلاط خبرگان می‌تواند به‌عنوان الگوریتم انتخاب دسته‌بند نیز دیده شود. دسته‌بندهای مجزا، خبرگانی در بخشی از فضای ویژگی هستند و قاعده ترکیب، مناسب‌ترین دسته‌بند را انتخاب می‌کند یا وزدهسته‌بندها بر اساس خبرگی‌شان برای دسته‌بندی نمونه آزمون تعیین می‌شود.

در پژوهش (Jordan and Jacobs, 1994)، روش پایه اختلاط خبرگان توسعه داده شده و «اختلاط خبرگان سلسله‌مراتبی»⁷ نامیده شده است. این روش، فضای مسأله را به زیرفضاهایی تجزیه می‌کند و به‌طور سلسله‌مراتبی هر زیرفضا را به زیرفضاهایی کوچک‌تر تقسیم می‌کند.

۲-۲- رویکرد زیرفضا: سطح ویژگی

یکی از رویکردهای مؤثر برای ایجاد شورایی از دسته‌بندهای گوناگون، استفاده از زیرمجموعه‌های ویژگی مختلف و آموزش دسته‌بندهای پایه با این زیرمجموعه‌هاست که «شورای انتخاب ویژگی»⁸ گفته می‌شود. هدف شورای انتخاب ویژگی، یافتن زیرمجموعه‌های ویژگی است که صحت دسته‌بندی خوبی داشته و تا حد ممکن گوناگون باشند.

زمانی که تعداد نمونه‌های مجموعه آموزش نسبت به ابعاد داده کم است، این رویکرد به‌طور معمول گزینه بهتری در مقایسه با روش‌های رویکرد زیرنمونه است (Skurichina and Duin, 2002). در این رویکرد، ابعاد زیرفضا کمتر از ابعاد فضای ویژگی اولیه خواهد بود؛ درحالی که تعداد نمونه‌های

دسته‌ای که بیشترین رأی (امتیاز) را از دسته‌بندها بگیرد، تصمیم شورای دسته‌بندها برای نمونه آزمون خواهد بود.

□ **Wagging**: این روش نسخه‌ای از الگوریتم bagging است که برای آموزش هر دسته‌بند، از همه نمونه‌ها استفاده می‌کند؛ اما وزن هر نمونه به‌طور تصادفی تعیین می‌شود (Bauer and Kohavi, 1999). درحقیقت bagging را می‌توان نوعی از الگوریتم wagging با توزیع وزنی پواسون دانست؛ چرا که هر نمونه چند بار (با مقادیر گسسته) در الگوریتم استفاده می‌شود.

□ **Rotation Forest**: این روش از دیگر روش‌های موفق ایجاد شورای دسته‌بندهاست که در سال ۲۰۰۶ میلادی مطرح شده است (Rodríguez, Kuncheva, and Alonso, 2006). در این روش با اعمال تحلیل مؤلفه‌های اصلی¹ بر روی زیرمجموعه آموزش هر دسته‌بند پایه، محورهای اولیه ویژگی می‌چرخد. ابتدا زیرمجموعه ویژگی F ، به‌طور تصادفی به K زیرمجموعه تقسیم و سپس PCA بر روی هر زیرمجموعه اعمال می‌شود. از این زیرمجموعه‌ها برای آموزش دسته‌بندهای پایه استفاده می‌شود. در این روش، تمام مؤلفه‌های اصلی حفظ می‌گردد تا پراکندگی اطلاعات داده حفظ شود. لذا K محور چرخش یافته به‌دست می‌آید تا ویژگی‌های جدیدی را برای دسته‌بندهای پایه بسازند. ایده اصلی این روش بهبود همزمان گوناگونی و صحت هر دسته‌بند در شورا است: گوناگونی از طریق استخراج ویژگی برای هر دسته‌بند پایه با استفاده از PCA و صحت با نگهداری تمام مؤلفه‌های اصلی و نیز استفاده از تمام مجموعه داده برای هر دسته‌بند به‌دست می‌آید. در پژوهش (Nanni and Lumini, 2009) توانایی این روش برای تشخیص دانش‌آموزان دارای مشکل ناتوانی یادگیری نشان داده شده؛ با این تفاوت که از «تحلیل مؤلفه‌های مستقل»² به جای PCA استفاده شده است.

□ **RotBoost**: این روش براساس ترکیب روش‌های Rotation Forest و AdaBoost ارائه شده است (Zhang and Zhang, 2008). آزمون این روش بر روی ۳۶ مجموعه داده UCI نشان داده که کارایی آن بهتر از روش‌های Rotation Forest، AdaBoost، MultiBoost و Bagging است.

□ **اختلاط خبرگان**³: در این روش که در سال ۱۹۹۱ میلادی مطرح شده است (Jacobs et al., 1991)، ابتدا دسته‌بندهای D_1, D_2, \dots, D_L شورای دسته‌بندها را تشکیل

⁴ gating network

⁵ gradient decent

⁶ Expectation Maximization (EM)

⁷ Hierarchical Mixtures of Experts (HME)

⁸ ensemble feature selection

¹ principal components analysis (PCA)

² independent component analysis (ICA)

³ mixture of experts

سپس آموزش دسته‌بندها با این مجموعه‌ها انجام می‌شود و سرانجام در مرحله آزمون، تصمیم اعضای شورا با قاعده ترکیب مناسبی ادغام می‌شوند.

در روش پیشنهادی «هو»، تعداد ویژگی‌های هر زیرمجموعه، m ، مقدار ثابت و به‌طور تقریبی برابر با نصف ویژگی‌های موجود، n ، است. به‌جای انتخاب تعدادی ثابت از ویژگی‌ها در هر زیرمجموعه، زیرمجموعه‌های ویژگی می‌تواند تعداد ویژگی‌های متفاوتی داشته باشد. با این تغییر، هر ویژگی بخت یکسانی برای انتخاب شدن در زیرمجموعه ویژگی دارد. همان‌گونه که پژوهش (Tsybmal and Puuronen, 2002) نشان می‌دهد این روش، گوناگونی شورای دسته‌بندها را افزایش می‌دهد و در نتیجه به کارایی بهتری در دسته‌بندی منجر می‌شود.

«بریل»⁷ و همکارانش روش Attribute Bagging (AB) را پیشنهاد کردند که مشابه رویکرد «هو» زیرمجموعه‌های تصادفی را با هم ترکیب می‌کند؛ با این تفاوت که در این روش زیرمجموعه‌های ویژگی بهتر با رویکرد پوششی⁸ انتخاب می‌شوند (Bryll, Gutierrez-Osuna, and Quek, 2003). در این روش، با جستجوی تصادفی ابعاد فضای ویژگی، اندازه مناسب زیرمجموعه‌ها تعیین می‌شوند. سپس زیرمجموعه‌های ویژگی مختلف به‌صورت تصادفی ایجاد شده و زیرمجموعه‌هایی را که نتایج بهتری بر روی نمونه‌های آموزشی داشتند برای دسته‌بندی نمونه‌های آزمون انتخاب می‌شوند. در پژوهشی دیگر (Tsybmal and Puuronen, 2002)، روشی برای ایجاد شورایی از دسته‌بندهای بیز با زیرمجموعه‌های تصادفی از ویژگی پیشنهاد شده است.

ب) انتخاب فروکاست‌های ویژگی: فروکاست کوچک‌ترین زیرمجموعه ویژگی است که توانایی پیش‌بینی یکسانی با مجموعه تمام ویژگی‌ها دارد. در این روش، اندازه شورای دسته‌بندها که با فروکاست‌های مختلف ویژگی ایجاد می‌شوند محدود به تعداد ویژگی‌هاست. پژوهش‌های مختلفی بر اساس این رویکرد انجام شده است. ایده ترکیب چندین دسته‌بند KNN برای دسته‌بندی متن با استفاده از فروکاست‌ها در (Bao and Ishii, 2002) ارائه شده است. روش «حذف بدترین ویژگی» برای یافتن مجموعه‌ای از فروکاست‌ها و ترکیب آنها با روش بیز ساده در (Wu and Bell, 2003) ارائه

آموزش ثابت می‌ماند. بنابراین اندازه نسبی مجموعه آموزش افزایش می‌یابد. همچنین زمانی که نمونه‌های مسأله دارای ویژگی‌های زائد زیادی است، استفاده از روش زیرفضای تصادفی نسبت به فضای اولیه نتایج بهتری به‌دست آید. تصمیم ادغامی این دسته‌بندها به‌طور معمول برتر از تصمیم دسته‌بند واحدی است که با مجموعه آموزش اولیه از تمام فضای ویژگی ساخته می‌شود. «هو»¹ نشان داد درحالی‌که اغلب روش‌های دسته‌بندی با مشکل ابعاد زیاد داده مواجه‌اند، روش زیرفضای تصادفی می‌تواند از این موضوع بهره‌مند شود (Ho, 1998).

نتایج پژوهش‌های اخیر نشان می‌دهد الگوریتم مناسب انتخاب زیرمجموعه‌های ویژگی، کارایی و صحت روش‌های زیرفضا را به‌طور قابل ملاحظه‌ای بهبود می‌دهد (Chen, Lin, and Chou, 2011)؛ چرا که کاهش فضای ویژگی «مشکل نمونه‌های با ابعاد زیاد»² را کاهش داده و فرایند آموزش و آزمون سریع‌تر انجام خواهد شد (Bryll, Gutierrez-Osuna, and Quek, 2003; Tumer and Oza, 2003). در الگوریتم پیشنهادی «هو»، زیرمجموعه‌های ویژگی به‌طور تصادفی ایجاد می‌شوند. با این حال انتخاب زیرمجموعه‌ها می‌تواند تصادفی نباشد. سه راهکار اصلی برای ایجاد شورای دسته‌بندها با زیرمجموعه‌های ویژگی مختلف می‌توان معرفی کرد (Rokach, 2008): الف) انتخاب تصادفی³ زیرمجموعه‌های ویژگی؛ ب) انتخاب فروکاست‌های⁴ ویژگی؛ و ج) انتخاب بر مبنای کارایی زیرمجموعه‌ها⁵.

الف) راه‌کار انتخاب تصادفی: ساده‌ترین روش ایجاد زیرمجموعه‌های ویژگی مختلف بر اساس انتخاب تصادفی ویژگی‌ها است. «هو» (Ho, 1998) جنگلی از درخت‌های تصمیم با انتخاب شبه‌تصادفی ویژگی‌ها ایجاد کرد و نشان داد انتخاب تصادفی زیرمجموعه‌های ویژگی می‌تواند روش مؤثرتری باشد؛ چرا که گوناگونی دسته‌بندها، کم‌بودن میزان صحت دسته‌بندی آنها را جبران می‌کند. این الگوریتم، «روش زیرفضای تصادفی»⁶ نامیده می‌شود. در این روش، تعداد معینی زیرمجموعه ویژگی m تایی به‌طور تصادفی از بردار ویژگی n بُعدی مسأله انتخاب می‌شود ($m < n$).

¹ Ho

² curse of dimensionality

³ random-based

⁴ reduct

⁵ performance-based strategy

⁶ Random Subspace Method (RSM)

⁷ Bryll

⁸ wrapper

که در ادامه آورده شده است. شایان ذکر است، هدف GA انتخاب بهترین زیرمجموعه ویژگی نیست؛ بلکه GA به دنبال یافتن دسته‌بندی‌هایی است که در مجموع از بهترین دسته‌بند پایه بهتر باشند. از این رو، دسته‌بندی‌های مجموعه، باید صحت دسته‌بندی قابل قبول و درعین حال گوناگون باشند.

رویکرد اول (Opitz 1999): (شکل ۶)، مدل‌سازی

رویکرد اول را برای مسأله‌ای با تعداد ده ویژگی نشان می‌دهد که جمعیت اولیه شامل چهار کروموزوم است. در این رویکرد، زیرمجموعه‌های ویژگی به طور مجزا انتخاب می‌شوند. هر کروموزوم نماینده یک زیرمجموعه ویژگی (یک دسته‌بند پایه) و مجموعه تمام کروموزوم‌های تولید شده، شورای دسته‌بندها را نشان می‌دهد. هر ژن می‌تواند مقادیر ۰ یا ۱ داشته باشد. مقدار ۱ برای ژن i ام ($1 \leq i \leq 10$) تعداد ویژگی (نشان دهنده انتخاب و مقدار ۰ نشان دهنده عدم انتخاب ویژگی) نام در آن زیرمجموعه است.

۰	۰	۱	۰	۱	۰	۰	۰	۱	۱	کروموزوم ۱
۰	۱	۰	۰	۱	۰	۰	۰	۰	۰	کروموزوم ۲
۱	۰	۰	۱	۰	۱	۰	۱	۱	۰	کروموزوم ۳
۰	۱	۱	۰	۰	۰	۱	۰	۰	۱	کروموزوم ۴

(شکل ۶): مدل‌سازی شورای انتخاب ویژگی با الگوریتم ژنتیک در این روش، گوناگونی دسته‌بندها به‌طور معمول به صورت فاصله زیرمجموعه‌های ویژگی محاسبه می‌شود. برای نمونه، فاصله کروموزوم اول از دیگر کروموزوم‌ها به ترتیب برابر ۴، ۷، و ۴ است. در نتیجه، میانگین گوناگونی دسته‌بند اول، برابر ۵ است. با این حال، محاسبه گوناگونی دسته‌بندها به این روش، گوناگونی خروجی دسته‌بندها را تضمین نمی‌کند. این معیار گوناگونی همراه با صحت گوناگونی هر دسته‌بند در تابع برازش شورا استفاده می‌شود.

رویکرد دوم (Kuncheva and Jain 2000): در این

رویکرد، هر کروموزوم در جمعیت، نمایانگر کل شورا است. همچنین، زیرمجموعه‌های ایجاد شده می‌توانند مجزا یا هم‌پوشان باشند. در شیوه مجزا، طول هر کروموزوم برابر با تعداد ویژگی‌های مسأله است. هر ژن می‌تواند مقادیر ۰ تا L (تعداد دسته‌بندی‌های شورا) را داشته باشد و نشان دهنده دسته‌بندی است که آن ویژگی را دارد. برای مثال، مسأله‌ای با تعداد ۱۰ ویژگی را در نظر بگیرید که شورا شامل ۴ دسته‌بند $\{D_1, D_2, D_3, D_4\}$ است. در این مسأله، کروموزوم به شکل $[۳, ۳, ۲, ۱, ۳, ۰, ۳, ۱, ۴, ۱]$ نشان می‌دهد که D_1 شامل ویژگی‌های ۱، ۳ و ۷؛ D_2 شامل ویژگی ۸، D_3 شامل ویژگی‌های ۴، ۶، ۹ و ۱۰؛ و D_4 شامل ویژگی ۲ است. ویژگی ۵ در هیچ زیرمجموعه‌ای استفاده نشده است. در

شده است. «هو» و همکارانش روشهای مختلفی را برای ساخت جنگل‌های تصمیم با فروکاست‌های متفاوت پیشنهاد کردند (Hu, Yu, and Wang 2005). همچنین ایجاد شورای دسته‌بندها با فروکاست‌های مختلف براساس نظریه مجموعه‌های شولا^۱ نیز نتایج بسیار خوبی را نشان می‌دهد (Hu et al. 2007; Wanga et al. 2010).

ج) انتخاب بر مبنای کارایی زیرمجموعه‌ها: در این

رویکرد، زیرمجموعه‌های ویژگی به گونه‌ای انتخاب می‌شوند که کارایی نهایی سیستم شورای دسته‌بندها افزایش یابد. در (Cunningham and Carney 2000) ابتدا دسته‌بندی‌های پایه با زیرمجموعه‌های تصادفی ویژگی ایجاد می‌شوند؛ سپس فرایند تکراری بهبود زیرمجموعه‌ها براساس روش جستجوی تپه‌نوردی^۲ انجام می‌شود تا صحت و گوناگونی دسته‌بندی‌های پایه افزایش یابد؛ به نحوی که برای تمام زیرمجموعه‌های ویژگی به دنبال کاستن یا افزودن یک ویژگی هستیم. اگر زیرمجموعه ویژگی جدید، کارایی بهتری بر روی نمونه‌های ارزیابی داشت، این تغییر انجام می‌شود. در غیر این صورت این فرآیند تا زمانی تکرار می‌شود که بهبود بیشتری به دست نیاید. در پژوهشی دیگر، روشی با نام «حذف یک‌در ده ورودی»^۳ (ID) ارائه شده که ویژگی‌ها براساس همبستگی آنها با برجسب دسته‌ها انتخاب می‌شوند (Oza and Tumer 2001) و نشان داده شده که این روش کارایی بهتری نسبت به انتخاب تصادفی زیرمجموعه‌های ویژگی دارد. استفاده از روش‌های مختلف انتخاب ویژگی همراه با معیار گوناگونی در تابع برازش برای انتخاب بهترین زیرمجموعه‌های ویژگی در (Tsybal, Pechenizkiy, and Cunningham 2005) بررسی شده است. این مقاله نشان می‌دهد برخی معیارهای گوناگونی به‌طور کلی نتایج بهتری می‌دهند.

همچنین روش‌های مختلفی بر مبنای الگوریتم ژنتیک برای یافتن زیرمجموعه‌های ویژگی ارائه شده است (Kuncheva and Jain 2000; Yu and Cho 2006; Opitz 1999). «کونچوا» دو رویکرد کلی برای ایجاد شورای انتخاب ویژگی با الگوریتم ژنتیک بیان می‌کند (Kuncheva 2004)

¹ rough sets

² hill-climbing search

³ input decimation (ID)

که در ابتدا توسط «دیتریچ»^۷ و «بکیری»^۸ ارائه شده است (Dietterich and Bakiri, 1995)، ماتریس M با ابعاد $N_c \times L$ ایجاد می‌کند، که L طول هر رشته کد است و هر عنصر این ماتریس مقادیر $\{1, -1\}$ را دارد. هر ستون این ماتریس نمایان‌گر یک دسته‌بند دودویی است که دودوساز^۹ نیز خوانده می‌شود. هر دودوساز، دسته‌های مسأله را به دو فرادسته^{۱۰} تقسیم می‌کند. اگر $M_{ij} = 1$ باشد، نمونه x که به دسته i تعلق دارد، نمونه مثبتی برای دسته‌بند i محسوب می‌شود و در غیر این صورت نمونه منفی خواهد بود. (شکل ۷)، یک نمونه ماتریس ECOC برای مسأله‌ای ۴ دسته‌ای را نشان می‌دهد که شامل شش دسته‌بند است. در این ماتریس، هر ستون متناظر با یک دسته‌بند و هر سطر متناظر رشته کدی منحصر به فرد برای هر دسته است. برای نمونه، دسته‌بند h_3 ، دو فرادسته را ایجاد می‌کند: دسته‌های ۱ و ۴ فرادسته اول و دسته‌های ۲ و ۳ فرادسته دوم را شکل می‌دهند. لذا دسته‌بند h_3 ، با نمونه‌های دسته ۱ و ۴ به عنوان نمونه‌های مثبت و نمونه‌های دسته ۲ و ۳ به عنوان نمونه‌های منفی آموزش می‌بیند.

dichotomizer

دسته						
۱	1	0	1	0	0	1
۲	1	1	0	1	1	0
۳	0	1	0	1	0	1
۴	1	0	1	1	1	1

(شکل ۷): مثالی از ماتریس ECOC برای مسأله‌ای با

۴ دسته و ۶ دسته‌بند

برای دسته‌بندی نمونه x خروجی هر دسته‌بند مقدار صفر یا یک خواهد بود؛ که در نتیجه بردار خروجی با طول L ایجاد خواهد شد. این بردار خروجی با رشته‌کدهای هر دسته مقایسه می‌شود و دسته‌ای که کمترین فاصله را با بردار خروجی دارد، به عنوان دسته نمونه آزمون پیش‌بینی می‌شود. فرآیند ادغام خروجی دسته‌بند های منفرد رمزگشایی^{۱۱} خوانده می‌شود. مرسوم‌ترین روش‌های رمزگشایی، تابع فاصله همینگ و نرم L^1 هستند که به دنبال کمترین فاصله بردار خروجی با هر رشته کد هستند:

$$y_H = \arg \min_{r \in \{1, \dots, N_c\}} \sum_{i=1}^L \left(\frac{1 - \text{sign}(M(r, i), f_i(x))}{2} \right),$$

⁷ Dietterich

⁸ Bakiri

⁹ dichotomizer

¹⁰ metaclass

¹¹ decoding

نشده است. در شیوه هم‌پوشان، هر شورای دسته‌بندی، با یک کروموزوم دودویی با طول $n \times L$ نمایش داده می‌شود. n بیت اول نمایان‌گر زیرمجموعه ویژگی اول برای D_1 است. به همین ترتیب n بیت‌های بعد زیرمجموعه‌های ویژگی دیگر را نمایش می‌دهند. در این روش، هر دسته‌بند می‌تواند ویژگی‌های مشترک با سایر دسته‌بندها را داشته باشد.

مشکل رویکرد اول این است که صحت دسته‌بندی شورا در فرآیند ارزیابی روش تأثیری ندارد. لذا، تضمینی وجود ندارد که شورا صحت دسته‌بندی بیشتری از بهترین دسته‌بند پایه (کروموزوم) داشته باشد. در مقابل، در رویکرد دوم، صحت شورای دسته‌بندی به طور دقیق محاسبه می‌شود. مزیت رویکرد اول، هزینه محاسباتی کمتر است.

۲-۳- سطح دسته‌بند

روش اصلی در ایجاد گوناگونی در سطح دسته‌بند، استفاده از دسته‌بند های مختلف مانند شبکه‌های عصبی، نزدیک‌ترین k همسایه، درخت‌های تصمیم و غیره است. رویکرد دیگر برای دستیابی به گوناگونی در این سطح، استفاده از پارامترهای آموزش متفاوت برای دسته‌بند های مختلف است. برای نمونه، دسته‌ای از شبکه‌های عصبی پرسپترون چند لایه^۱ می‌تواند با وزن‌های اولیه متفاوت، تعداد لایه‌ها و نرون‌های متفاوت، خطای هدف متفاوت و ... آموزش ببیند. تنظیم این پارامترها کنترل ناپایداری^۲ دسته‌بند های مجزا را ممکن می‌سازد و لذا به گوناگونی آنها منجر می‌شود. توانایی کنترل ناپایداری شبکه‌های عصبی و دسته‌بند هایی از نوع درخت تصمیم، آنها را نامزدهایی مناسب برای استفاده در سیستم دسته‌بند شورایی می‌کند. از سوی دیگر، دسته‌بند های به‌طور کامل متفاوت، مانند شبکه‌های عصبی MLP، درخت‌های تصمیم^۳، دسته‌بند های نزدیک‌ترین همسایه، و ماشین‌های بردار پشتیبان^۴ می‌توانند برای افزودن گوناگونی ترکیب شوند (Kuncheva, 2004).

۲-۴- کدهای خروجی تصحیح‌کننده خطا^۵

از نگاه سیستم‌های شورایی، روش ECOC را می‌توان از جمله رویکردهای شورایی مناسب برای حل مسائل چنددسته‌ای دانست. پایه روش ECOC طراحی یک رشته کد^۶ برای هر کدام از دسته‌های مسأله است. این روش

¹ multilayer perceptron (MLP)

² instability

³ decision trees

⁴ support vector machines

⁵ Error Correcting Output Codes (ECOC)

⁶ codeword

روش‌های افراز فضای ویژگی، کارایی بهتری نسبت به روش‌های زیرنمونه (مانند bagging و boosting) نشان می‌دهند (Banfield et al., 2007; Tumer and Ghosh, August 1996).

برخی پژوهش‌ها روش‌های مهم شورای دسته‌بندها را مقایسه کرده‌اند. مقایسه روش‌های ایجاد شورا با دسته‌بند درخت تصمیم در (Banfield et al., 2007)، از جامع‌ترین پژوهش‌های اخیر در سیستم‌های شورای است. در این مقاله، روش bagging با روش‌های RSM، دو نسخه از روش AdaBoost (شامل ۵۰ و ۱۰۰۰ درخت)، و سه نسخه از روش جنگل تصادفی (RF) با تعداد یک، دو و لگاریتم نصف به‌علاوه یک ویژگی (RF.lg) مقایسه شده است. نتایج تجربی این پژوهش با ۵۷ مجموعه داده نشان می‌دهد روش AdaBoost.1000 و RF.lg بهترین صحت دسته‌بندی را نتیجه می‌دهند. با این حال، برای ۳۷ مجموعه داده آزمایش، هیچ کدام از این روش‌های موجود تفاوت معناداری را در مقایسه با bagging نشان نمی‌دهند. مقایسه روش‌های طراحی شورا شامل bagging، AdaBoost، Arc-X4، دسته‌بندهای SVM نتایج متفاوتی را نشان می‌دهد (Wang et al., 2009). بر اساس نتایج تجربی این پژوهش با استفاده از بیست مجموعه داده، bagging-SVM به‌ویژه با تابع هسته چندجمله‌ای^۴ صحت دسته‌بندی بهتری را ارائه می‌کند. یکی از نتایج بسیار مهم این پژوهش این است که در بسیاری از مجموعه داده‌های آزمایش، روش‌های شورای با استفاده از SVM، صحت دسته‌بندی بدتری را در مقایسه با دسته‌بند منفرد SVM نتیجه می‌دهند. برای نمونه، میانگین صحت SVM با هسته تابع پایه شعاعی، از تمام روش‌های شورای بهتر است. این نتایج تا حد زیادی با نتایج پژوهش (Meyer, Leisch, and Hornik, 2003) مشابه است.

۲-۵- راهکار آزمون-انتخاب^۵

هرچند روش‌های مختلفی برای تشکیل شورای دسته‌بندها وجود دارد، تعیین بهترین روش برای مسأله خاص موضوعی فراخور پژوهش است. برای این موضوع، رویکرد «آزمون و انتخاب» (Sharkey et al., 2000) به‌عنوان گزینه‌ای مناسب ارائه شده است؛ هرچند که از «جواب اصلی»^۶ دور باشد. مرحله اول این روش، فراتولید^۷ ترکیب‌های ممکن

$$y_{L^1} = \arg \min_{r \in \{1, \dots, N_c\}} \sum_{i=1}^L |f_i(x) - M(r, i)|$$

که اگر $z > 0$ ، $sign(z)$ برابر با ۱؛ اگر $z < 0$ ، $sign(z)$ برابر با -۱؛ و در غیر این صورت صفر خواهد بود. $M(r, \cdot)$ نشان‌دهنده رشته کد r در ماتریس است و $y \in \{1, \dots, N_c\}$ دسته پیش‌بینی نمونه آزمون را نشان می‌دهد. روش‌های رمزگشایی مختلفی علاوه بر روش‌های فاصله ارائه شده است که برای مطالعه بیشتر مرجع (Windeatt and Ghaderi, 2003) پیشنهاد می‌شود.

روش ECOC توسط «آلوین»^۱ و همکارانش توسعه داده شد (Allwein, Schapire, and Singer, 2001). در این شیوه، عناصر ماتریس ECOC می‌توانند شامل مقادیر $\{1, 0, -1\}$ باشند؛ که مقدار صفر نشان‌دهنده عدم استفاده از یک دسته در فرآیند آموزش دسته‌بند است. در (Allwein, Schapire, and Singer, 2001)، هر عنصر ماتریس مقدار صفر با احتمال $0.5/5$ و مقدار $+1$ یا -1 با احتمال $0.25/5$ دارد. از این‌رو، نمونه‌های یک دسته می‌توانند در فرآیند آموزش یک دسته‌بند نادیده گرفته شوند. این روش ایجاد ماتریس ECOC، کدهای تصادفی تنک^۲ خوانده شده؛ در حالی که روش اولیه ایجاد ماتریس ECOC کدهای تصادفی متراکم^۳ گفته می‌شود. با استفاده از این رویکرد یکپارچه، روش‌های قدیمی مسائل چنددسته‌ای (یک-درمقابل-یک و یک-درمقابل-همه) را می‌توان به‌آسانی در قالب ماتریس ECOC بازنمایی کرد. در روش یک-درمقابل-یک، ماتریس ECOC شامل $\binom{N_c}{2}$ ستون است؛ به طوری که هر ستون متناظر با یک زوج دسته ω_i, ω_j است. در این ستون، مقدار سطر ω_i برابر $+1$ ، مقدار سطر ω_j برابر -1 و مقدار بقیه عناصر صفر است. در رویکرد یک-درمقابل-همه، M ماتریسی با ابعاد $N_c \times N_c$ است که مقادیر عناصر قطری آن برابر $+1$ و مقادیر سایر عناصر -1 است.

بحث و نتیجه‌گیری: هرچند روش‌های معرفی شده

بر اساس رویکرد زیرنمونه، کارایی خود را برای مجموعه داده‌های مختلف نشان داده‌اند، اما محبوبیت روش‌های Bagging و AdaBoost در حوزه‌های مختلف کاربردی به مراتب بیشتر است. همچنین نتایج برخی پژوهش‌های اخیر نشان می‌دهد که آموزش دسته‌بندها با زیرمجموعه‌های ویژگی مختلف یکی از مؤثرترین روش‌های ایجاد گوناگونی در شورای دسته‌بندهاست. به بیان دیگر،

⁴ polynomial kernel

⁵ test-select strategy

⁶ complete solution

⁷ over-producing

¹ Allwein

² sparse random codes

³ dense random codes

یادگیری دیگری تعیین می‌شود. به این روش‌ها روش‌های آموزش‌پذیر^۶ گفته می‌شود.

۳-۲- انواع خروجی دسته‌بندها

خروجی الگوریتم‌های مختلف دسته‌بندی به یکی از شکل‌های زیر بیان می‌شود (Xu, Krzyzak, and Suen, 1992):

الف. خروجی تک‌کلاسه^۷ (مطلق): در این شکل، دسته‌بند فقط یک برچسب برای الگوی ورودی (x) ارائه می‌کند. در این نوع، اطلاعاتی در خصوص قطعیت برچسب پیشنهادی وجود ندارد. به‌طور معمول، تمامی دسته‌بندها می‌توانند یک برچسب را برای الگوریتم ورودی تعیین کنند؛ لذا خروجی مطلق بیشترین کاربرد را دارد.

ب. خروجی رتبه‌ای^۸: در این حالت، دسته‌بند تعلق الگوی ورودی به دسته‌های مختلف را به‌صورت فهرستی مرتب‌شده ارائه می‌کند. دسته با بیشترین تعلق در ابتدای فهرست و دسته با کمترین تعلق در انتهای آن قرار دارد. خروجی رتبه‌ای به‌ویژه برای مسائل با تعداد دسته‌های زیاد مانند تشخیص حرف، صورت و صدا مناسب است.

ج. خروجی امتیازدار^۹ (پیوسته): در این حالت، دسته‌بند برای تعلق الگوی ورودی به هر کدام از دسته‌ها یک مقدار عددی تخصیص می‌دهد. به‌عبارتی خروجی هر دسته‌بند D_i برداری c -بعدی $[d_{i,1}, \dots, d_{i,c}]$ خواهد بود؛ مقدار $d_{i,j}$ نشان‌دهنده امتیازی^{۱۰} است که نمونه x متعلق به دسته j -ام است.

واضح است که خروجی تک‌کلاسه شامل کمترین اطلاعات و خروجی امتیازدار حاوی بیشترین اطلاعات در فرایند دسته‌بندی است. متداول‌ترین تقسیم‌بندی برای روش‌های ترکیب خروجی دسته‌بندها، تقسیم‌بندی آنها بر حسب نوع خروجی الگوریتم دسته‌بندی است که بر اساس آن روش‌های ترکیب به سه بخش تقسیم می‌شوند. (شکل ۸) روش‌های مختلف ترکیب دسته‌بندها را به‌خوبی نشان می‌دهد. بر اساس این تقسیم‌بندی، روش‌های مهم هر بخش در ادامه آورده شده است.

شوراهاست. سپس این شوراها با مجموعه اعتباریابی^۱ محک زده می‌شوند و بهترین شورا انتخاب می‌شود. درنهایت، این شورا با مجموعه آزمون (که در فرایند انتخاب وارد نشده بود)، آزمایش می‌شود. از این رو رویکرد «آزمون و انتخاب» در عمل به دو مجموعه آزمون نیاز دارد: مجموعه اعتباریابی و مجموعه آزمون؛ که به ترتیب برای انتخاب شورا و آزمون کارایی آن استفاده می‌شوند. مجموعه اعتباریابی، می‌تواند مانع از فراتخمین^۲ کارایی شورا شود که به‌طور معمول زمانی که انتخاب و ارزیابی براساس یک مجموعه آزمون است، رخ می‌دهد.

۳- طراحی قواعد ترکیب

موضوع مهم دیگر در ساخت شورای دسته‌بندها، انتخاب قاعده ترکیب دسته‌بندهای پایه است. انتخاب قاعده ترکیب، که تحت عنوان بهینه‌سازی تصمیم نیز نامیده می‌شود، ساخت تابع ترکیب قدرتمندی برای ادغام تصمیم دسته‌بندهای مجزاست. به بیان دیگر بهینه‌سازی تصمیم سعی در یافتن قاعده ترکیب بهینه‌ای دارد که تصمیم مجموعه دسته‌بندها را ادغام کند. نشان داده شده که در صورت عدم استفاده از روش ترکیب مناسب، مزیت گوناگونی دسته‌بندها موجب بهبود کارایی شورا نخواهد شد (Brodley and Lane 1996). در ادامه پس از معرفی دو مفهوم مهم، مهم‌ترین قواعد ادغام دسته‌بندها آورده شده است. در این مقاله، روش‌هایی که در مقاله (نبوی کریزی و کبیر ۱۳۸۴) معرفی شده است، به‌اجمال بیان شده است.

۳-۱- روش‌های ترکیب آموزش‌پذیر و غیر آموزش‌پذیر

ترکیب غیر آموزش‌پذیر^۳ به این معناست که روش ترکیب‌کننده، پارامتر اضافه‌ای برای آموزش ندارد؛ به بیان دیگر، پس از اینکه آموزش دسته‌بندهای پایه تمام شد، می‌توان از شورا برای آزمون نمونه‌ها استفاده کرد. نمونه‌ای از این نوع قواعد ترکیب، قاعده رأی حداکثر^۴ و شمارش بورد^۵ است. در برخی از روش‌های ترکیب، علاوه بر آموزش دسته‌بندهای پایه، پارامترهای روش نیز از طریق الگوریتم

⁶ trainable

⁷ abstract level

⁸ rank level

⁹ measurement level

¹⁰ support

¹ validation

² over-estimation

³ non-trainable

⁴ majority vote

⁵ Borda count

زیرمجموعه دسته‌بندها $A = \{D_{i,1}, \dots, D_{i,k}\}$, $A \in D$

احتمال توأم آنها قابل تجزیه به صورت زیر است:

$$P(D_{i_1} = s_{i_1}, \dots, D_{i_k} = s_{i_k}) = P(D_{i_1} = s_{i_1}) \times \dots \times P(D_{i_k} = s_{i_k}) \quad (\Delta)$$

$$P(D_{i_1} = s_{i_1}, \dots, D_{i_k} = s_{i_k}) = P(D_{i_1} = s_{i_1}) \times \dots \times P(D_{i_k} = s_{i_k})$$

که برچسب خروجی دسته‌بند $D_{i,j}$ است.

اگر حداقل $[L/2] + 1$ دسته‌بند تصمیم صحیح

بگیرند، رأی اکثریت دسته صحیح را انتخاب می‌کند. لذا

احتمال دسته‌بندی صحیح شورا برابر است با:

$$P_{maj} = \sum_{m=[L/2]+1}^L \binom{L}{m} p^m (1-p)^{L-m}$$

$$P_{maj} = \sum_{m=[L/2]+1}^L \binom{L}{m} p^m (1-p)^{L-m} \quad (6)$$

این احتمال به‌ازای $p = 0.6, 0.7, 0.8, 0.9$

در جدول ۲ نشان داده شده است. این نتایج

«نظریه هیئت منصفه کندرست»^۴ نیز خوانده می‌شود:

(جدول ۱): مقادیر صحت دسته‌بندی به‌ازای مقادیر مختلف L دسته‌بند

مستقل و صحت دسته‌بندی p

	$L=3$	$L=5$	$L=7$	$L=9$
$p=0.6$	0.6480	0.6826	0.7102	0.7334
$p=0.7$	0.7840	0.8369	0.8740	0.9012
$p=0.8$	0.8960	0.9421	0.9667	0.9804
$p=0.9$	0.9720	0.9914	0.9973	0.9991

از رابطه فوق چند نکته نتیجه می‌شود:

۱. اگر $p > 0.5$ ، آنگاه P_{maj} در رابطه ۶ به‌طور یکنواخت

$$L \rightarrow \infty : P_{maj} \rightarrow 1$$

افزایش می‌یابد؛ آنگاه P_{maj} در رابطه ۶ به‌طور یکنواخت

$$L \rightarrow \infty : P_{maj} \rightarrow 0$$

کاهش می‌یابد؛ اگر $p = 0.5$ ، آنگاه به‌ازای هر تعداد دسته‌بند:

$$P_{maj} = 0.5$$

این نتایج نشان می‌دهد، تنها زمانی که صحت

دسته‌بندی p بیش از 0.5 است، می‌توان انتظار بهبود

دسته‌بندی شورا را داشت. در این حالت، با شرط مستقل

بودن خطای دسته‌بند صحت شورای دسته‌بند می‌تواند

به ۱ میل کند.

۳-۳-۲- رأی‌گیری حداکثری وزن‌دار

اگر دسته‌بندها در شورا کارایی یکسانی نداشته باشند،

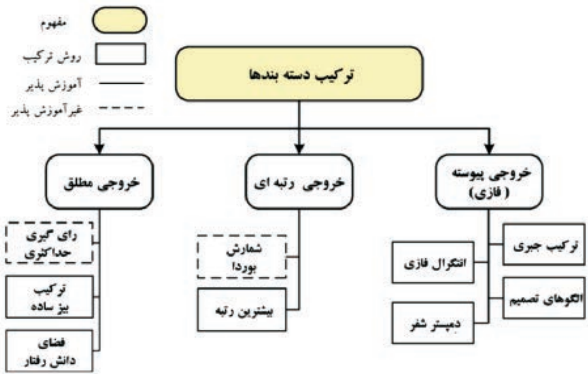
منطقی است که برای تصمیم‌گیری شورا به دسته‌بندهای

برتر وزن بیشتری داده شود. مشابه پیش، فرض کنیم تصمیم

دسته‌بند D_i برای دسته ω_j باشد؛ به‌طوری که اگر D_i

^۳ [۱] جزء صحیح عدد را نشان می‌دهد

^۴ Condorcet Jury Theorem (1785)



(شکل ۸): روش‌های مختلف ترکیب دسته‌بندها، برگرفته از

(Ruta and Gabrys, 2000)

۳-۳-۳- روش‌های خروجی مطلق (خروجی

تک‌دسته‌ای)

۳-۳-۱- رأی اکثریت

رأی اکثریت ساده‌ترین روش ترکیب خروجی دسته‌بندها

است. در این روش، خروجی نمونه آزمون دسته‌ای است که

بیشترین رأی را در میان دسته‌بندها داراست. زمانی که

مسئله تنها شامل دو دسته است، به این روش رأی‌گیری

حداکثری ساده^۱ نیز گفته می‌شود. در این حالت باید حداقل

$50\% + 1$ دسته‌بندها بر یک دسته خاص توافق داشته باشند.

توافق کامل^۲ از دیگر روش‌های مرسوم در این حوزه است. در

این روش، تنها زمانی که تمامی دسته‌بندها دسته مشترکی

برای نمونه آزمون تعیین کنند، تصمیم گرفته می‌شود. این

سه رویکرد در (شکل ۹) نشان داده شده است (Kuncheva,

1997; Lam and Suen., 2004).

توافق کامل

حداکثری ساده (50% + 1)

رأی‌گیری حداکثری

(شکل ۹): الگوهای اتفاق آرا در میان ۱۰ تصمیم‌گیر

(در تمام این سه روش تصمیم‌گیری نهایی گروه، مشکی است)

برای درک علت محبوبیت روش رأی‌گیری حداکثری

فرض کنید:

- تعداد دسته‌بندها، L ، فرد است.

- احتمال اینکه هر دسته‌بند دسته صحیح را برای هر

$x \in R^n$ انتخاب کند برابر p است.

- خروجی دسته‌بندها مستقل هستند؛ یعنی برای هر

^۱ simple majority

^۲ unanimity

برای هر دسته‌بند D_i ، ماتریس ابهام^۲ (CM^i) با ابعاد $c \times c$ با استفاده از نمونه‌های آموزش تشکیل می‌شود. عنصر (k, s) ام این ماتریس $(CM_{k,s}^i)$ ، بیانگر تعداد نمونه‌های در مجموعه آموزش است که برچسب واقعی آنها k است و توسط دسته‌بند D_i به دسته s تخصیص داده شده‌اند. با جمع کردن مقادیر عناصر ستون s ام ماتریس ابهام این دسته‌بند، می‌توان $CM_{.,s}^i$ (یعنی تعداد کل عناصری که این دسته‌بند به دسته s ω_s نسبت داده است) را مشخص کرد:

$$CM_{.,s}^i = \sum_{c=1}^c CM_{k,s}^i \quad (9)$$

با استفاده از مقادیر $CM_{.,s}^i$ و $CM_{k,s}^i$ ، می‌توان ماتریس برچسب^۴ LM^i ، را محاسبه کرد. این ماتریس، ماتریسی با ابعاد $c \times c$ است که عنصر (k, s) ام آن، تخمینی از احتمال آن است که دسته واقعی k باشد، به شرطی که دسته‌بند D_i دسته قطعی s را تخصیص داده باشد (رابطه ۱۰):

$$LM_{.,s}^i = \hat{P}(\omega_k | D_i(x) = \omega_s) = \frac{CM_{k,s}^i}{CM_{.,s}^i} \quad (10)$$

اگر S_1, S_2, \dots, S_L برچسب‌هایی باشد که دسته‌بندهای D_1 تا D_L به الگوی x تخصیص داده‌اند، با فرض استقلال دسته‌بندها، تخمین احتمال اینکه برچسب واقعی الگوی x ، ω_s باشد از رابطه ۵ به دست می‌آید:

$$\mu_D^c(x) = \prod_{i=1}^L \hat{P}(\omega_k | D_i(x) = S_i) = \prod_{i=1}^L LM_{c,S_i}^i \quad i = 1, \dots, c \quad (11)$$

که در آن $\hat{P}(\omega_k | D_i(x) = S_i)$ با توجه به نمونه‌های آموزشی تخمین زده می‌شوند. برای این کار، تعداد الگوهای که توسط D_i برچسب S_i خورده‌اند ولی برچسب واقعی آنها ω_c است بر تعداد کل الگوهای نسبت داده شده توسط D_i به دسته ω_i تقسیم می‌شوند. توضیحات بیشتر این روش در (سیدحسین نبوی کریزی احسان اله کبیری پائیز ۱۳۸۴؛ Kuncheva, 2004) آورده شده است.

۳-۳-۴- فضای دانش رفتار^۵

بیشتر روش‌های ترکیب از فرض استقلال دسته‌بندها برای تصمیم‌گیری نهایی استفاده می‌کنند. این فرض همواره صحیح نیست. روش فضای دانش رفتار، BKS به چنین فرضی برای تصمیم‌گیری نهایی نیاز ندارد. این روش که توسط «هانگ»^۶ و «سوئن»^۷ (Huang and Suen, 1993) ارائه شده

دسته ω_j را انتخاب کند d_{ij} مقدار ۱ و در غیر این صورت ۰ است. همچنین فرض می‌کنیم روشی برای اندازه‌گیری کارایی هر دسته‌بند وجود دارد و براساس این کارایی به هر دسته‌بند وزن w_i می‌دهیم. لذا تصمیم شورا برای انتخاب دسته j ام بر اساس رابطه زیر است:

$$\sum_{i=1}^L w_i d_{i,j} = \max_{j=1:c} \sum_{i=1}^L w_i j \quad (7)$$

یعنی، دسته‌ای که بیشترین امتیاز وزن دار را بگیرد، تصمیم شورا خواهد بود. برای تفسیر راحت‌تر، می‌توان وزن‌ها را هنجار کرد؛ به طوری که مجموع آنها یک شود (Polikar, 2006).

پرسش اصلی این است که «چگونه وزن‌ها تعیین شوند؟». به طور مشخص اگر می‌دانستیم کدام دسته‌بندها بهتر عمل می‌کنند، بیشترین وزن‌ها را به آنها می‌دادیم یا شاید تنها آن دسته‌بندها را انتخاب می‌کردیم. زمانی که این دانش را نداریم، راهکار معقول استفاده از کارایی دسته‌بند بر روی مجموعه داده‌های اعتباریابی یا حتی بر روی داده‌های آموزش، به عنوان تخمینی از کارایی آتی دسته‌بند، است. این راهکار در الگوریتم AdaBoost (روابط (۲) و (۴)) استفاده شده است: AdaBoost وزنی برابر $\ln\left(\frac{1}{\beta_k}\right)$ به هر دسته‌بند تخصیص می‌دهد که $\beta_k = \frac{\epsilon_k}{1-\epsilon_k}$ و ϵ_k خطای وزن دار دسته‌بند D_k است. همانگونه که بیان کردیم، هنجارسازی خطا این امکان را می‌دهد که از معیار خطای β استفاده کنیم که در بازه $[0, 1]$ قرار دارد (به جای ϵ که بین 0 و $1/2$ است). اما دلیل مهم دیگر برای هنجارسازی این است که می‌توان نشان داد که اگر L دسته‌بند مستقل وجود داشته باشند، صحت شورای دسته‌بندها زمانی حداکثر می‌شود که وزن دسته‌بندها تناسب زیر را داشته باشند (Kuncheva, 2004):

$$w_k \propto \log \frac{p_k}{1-p_k} \quad (8)$$

از آنجا که صحت هر دسته‌بند $p_k = 1 - \epsilon_k$ است، مشاهده می‌شود که در الگوریتم AdaBoost وزن هر دسته‌بند به طور دقیق براساس رابطه ۸ تخصیص داده شده است.

۳-۳-۳- ترکیب بیز ساده^۱

این روش فرض می‌کند که دسته‌بندها مستقل از یکدیگرند و لذا بیز ساده نیز نامیده می‌شود. در این روش،

¹ naïve Bayes

² confusion matrix

⁴ label matrix

⁵ Behavior Knowledge Space

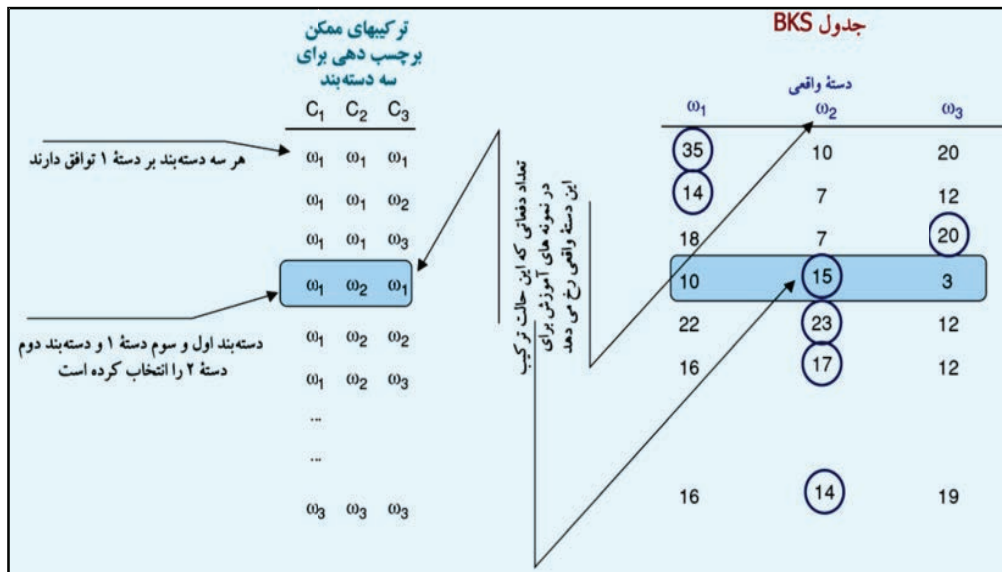
⁶ Huang

⁷ Suen

برای این حالت، اگر ترکیب دسته‌بندها با روش رأی‌گیری حداکثری باشد ω_1 انتخاب خواهد شد. اگر آموزش دسته‌بندها با تعداد زیاد نمونه‌های آموزش انجام شود و جدول به روش مناسبی هنجار شود، رأی حداکثر با استفاده از جدول BKS می‌تواند احتمال‌های پسین دسته را دقت خوبی تخمین بزند.

است، از جدول جستجو^۱ برای تعیین برچسب نمونه‌آزمون استفاده می‌کند (Huang and Suen, 1995).

با ثبت نظر دسته‌بندها در مورد الگوهایی که دسته آنها معلوم است، روش BKS رفتار جمعی دسته‌بندها را مدل کرده و بر اساس این مدل برچسب دسته الگوی ناشناخته x را مشخص می‌کند. به عبارت دیگر، در این روش تصمیم‌گیری نهایی برای دسته یک الگو با استفاده از رفتاری که دسته‌بندهای مختلف در هنگام یادگیری الگوها از خود نشان داده‌اند صورت می‌گیرد.



(شکل ۱۰): نمایش روش فضای دانش رفتار؛ برگرفته از (Polikar 2006)

اگر L دسته‌بند و C کلاس داشته باشیم، با احتساب دسته رد^۲ برای الگوها، جدول BKS $(C + 1)^L$ مقدار خواهد داشت. برای محاسبه جدول BKS به مجموعه نمونه‌های زیادی نیاز است. به عنوان نمونه اگر برای هر حالت ترکیب برچسب‌ها به ده نمونه نیاز باشد، برای حالت پنج دسته‌بند و سه دسته الگو، به تعداد 10×24 نمونه آموزشی $(10 \times (C + 1)^L = 10 \times (3 + 1)^5 = 1024)$ نیاز خواهد بود. هر چه تعداد داده بیشتر باشد، دقت روش بیشتر خواهد بود. واضح است که ترکیب با این روش زمانی موفق خواهد بود که دسته‌بندها، الگوها را تحت تمامی شرایط ممکن یاد گرفته باشند.

در (شکل ۱۰)، این روش با استفاده از یک مثال نشان داده شده است. فرض کنید سه دسته‌بند D_1, D_2, D_3 داریم و مسأله شامل سه دسته است. ۲۷ ترکیب ممکن برای برچسب‌دهی یک الگو وجود دارد که از $\{\omega_1 \omega_1 \omega_1\}$ تا $\{\omega_3 \omega_3 \omega_3\}$ مرتب شده‌اند. در طول فرایند آموزش، سابقه دادن هر ترکیب و نیز دسته صحیح متناظر با آن ترکیب را ثبت می‌کنیم. تعداد پیشامدهای هر ترکیب و هر دسته آورده شده است و حداکثر آن با دایره مشخص شده است. برای مثال، فرض کنید ترکیب $\{\omega_1 \omega_2 \omega_1\}$ در طول فرایند آموزش در مجموع ۲۸ بار رخ داده است؛ ۱۰ بار از آن دسته صحیح ω_1 ، در ۱۵ بار از آن دسته صحیح ω_2 و در ۳ بار از آن دسته صحیح ω_3 بوده است. لذا برنده این حالت ترکیب، ω_2 است: دسته‌ای که بیشترین مشاهده صحیح را برای این حالت ترکیب برچسب‌ها داشته است. بدین ترتیب، در فرایند آزمون هر زمان که ترکیب $\{\omega_1 \omega_2 \omega_1\}$ رخ دهد، این روش دسته ω_2 را انتخاب می‌کند. قابل توجه است که

۳-۴- روش‌های ترکیب برای خروجی رتبه‌ای

۳-۴-۱- شمارش بوردا

شمارش بوردا از یک نگاه مهم، با قواعدی که تاکنون معرفی شده متفاوت است: این روش، دسته‌های بازنده^۳ را نادیده

^۲ reject

^۳ دسته‌ای که دسته‌بند از دسته‌بندی نمونه‌های آن ناتوان است

^۱ look up table

$DP(x)$ ، مرتب شود (Kuncheva, Bezdek, and Duin, 2001):

$$DP(x) = \begin{bmatrix} d_{1,1}(x) & \dots & d_{1,j}(x) & \dots & d_{1,c}(x) \\ \vdots & & \vdots & & \vdots \\ d_{i,1}(x) & \dots & d_{i,j}(x) & \dots & d_{i,c}(x) \\ \vdots & & \vdots & & \vdots \\ d_{L,1}(x) & \dots & d_{L,j}(x) & \dots & d_{L,c}(x) \end{bmatrix} \quad (16)$$

سطر i ام ماتریس، خروجی دسته‌بند D_i برای الگوی x و ستون j ام ماتریس، بردار تعلق الگوی x به دسته w_j است که به کمک دسته‌بند D_1 تا D_L ارائه شده است. مقدار $d_{i,j}(x)$ با توجه به مقدار شباهت الگو به دسته‌های مختلف مقداری بین صفر و یک است. شایان ذکر است ماتریس پروفایل تصمیم برای دسته‌بند D_i با خروجی مطلق و خروجی رتبه‌ای نیز می‌تواند به کار گرفته شود. در دسته‌بند D_i با خروجی مطلق (تک‌دسته‌ای)، مقدار $d_{i,j}(x)$ صفر یا یک و برای دسته‌بند D_i با خروجی رتبه‌ای، $d_{i,j}(x)$ یک عدد صحیح است که رتبه تعلق الگوی x به دسته w_j را نشان می‌دهد.

۳-۵-۱- روش‌های ترکیب جبری

به‌طور کلی روش‌های ترکیب جبری ساده، ترکیب‌کننده‌هایی «غیرآموزش‌پذیر» برای خروجی‌های پیوسته هستند. امتیاز کلی هر دسته با استفاده از تابعی ساده از خروجی دسته‌بند w_j به دست می‌آید. قواعد ترکیب ساده امتیاز هر دسته w_j را تنها با استفاده از ستون i ام ماتریس $DP(x)$ محاسبه می‌کنند:

$$\mu_j(x) = \mathcal{F}[d_{1,j}(x), \dots, d_{L,j}(x)] \quad (17)$$

که \mathcal{F} تابع ترکیب است. دسته الگوی x ، اندیس بیشترین $\mu_j(x)$ است. تابع ترکیب \mathcal{F} می‌تواند به روش‌های مختلفی تعیین شود. مرسوم‌ترین این روش‌ها عبارتند از (Kuncheva, Bezdek, and Duin 2001):

□ میانگین ساده ($\mathcal{F} = \text{Average}$)

$$\mu_j(x) = \frac{1}{L} \sum_{i=1}^L d_{i,j}(x) \quad (18)$$

قاعده میانگین معادل با قاعده مجموع (همراه با ضریب هنجارسازی $\frac{1}{L}$) است. در هر دو روش، تصمیم شورا دسته‌ای خواهد بود که امتیاز $\mu_j(x)$ آن بیشترین مقدار است.

□ کمینه / بیشینه / میانه: این روش‌ها، مقدار کمینه، بیشینه یا میانه را از میان خروجی دسته‌بند انتخاب می‌کند:

نمی‌گیرد. در این روش برای هر دسته c ، «شمارش بوردا» برابر با مجموع تعداد دسته‌هایی که رتبه آنها پایین‌تر از دسته c است. اگر $B_{k,c}(x)$ تعداد دسته‌هایی باشد که برای نمونه آزمون x توسط دسته‌بند k ام پایین‌تر از دسته c مرتب شده‌اند، تصمیم نهایی شورا دسته‌ای است که بیشترین شمارش بوردا را دارد (Wanas and Kamel 2001):

$$\mu(x) = \operatorname{argmax}_c \sum_{k=1}^L B_{k,c}(x) \quad (12)$$

پیاپی این روش ساده است و نیازمند آموزش نیست. با این حال، این روش بین کارایی دسته‌بندها تفاوت قائل نمی‌شود و همه دسته‌بند یکسان در نظر گرفته می‌شوند؛ لذا زمانی که می‌دانیم برخی دسته‌بندها به احتمال برتر هستند این روش ترجیح داده نمی‌شود. برای مسأله با دو دسته، روش شمارش بوردا معادل روش رأی حداکثر است. روش شمارش بوردا به‌طور معمول زمانی که خروجی دسته‌بند به شکل رتبه‌ای است استفاده می‌شود. زمانی که خروجی دسته‌بند از نوع پیوسته است، می‌توان مقادیر خروجی را رتبه‌بندی کرده و از این روش برای ترکیب خروجی دسته‌بندها استفاده کرد.

۳-۵-۲- ترکیب خروجی‌های امتیازدار

فرض کنید \mathcal{D} و Ω به ترتیب بیان‌گر مجموعه دسته‌بندها و مجموعه دسته‌ها باشند:

$$\mathcal{D} = \{D_1, \dots, D_L\}, \quad \Omega = \{\omega_1, \omega_2, \dots, \omega_c\} \quad (13)$$

هر دسته‌بند D_i به‌ازای بردار ویژگی $x \in \mathcal{R}^n$ یک بردار خروجی به صورت:

$$D_i(x) = [d_{i,1}(x) \quad d_{i,2}(x) \quad \dots \quad d_{i,c}(x)] \quad (14)$$

تولید می‌کند که در آن $d_{i,j}(x)$ امتیاز تعلق الگوی x به دسته w_j است که توسط دسته‌بند D_i گزارش شده است. ادغام خروجی دسته‌بندها به این معناست که با ترکیب بردارهای خروجی L دسته‌بند، بردار تعلق الگوی x به دسته‌های c گانه را به صورت $D(x) = F(D_1(x), \dots, D_L(x))^T$ تعیین کنیم که در آن F قاعده ترکیب است. اگر $D(x)$ به صورت زیر فرض شود:

$$D(x) = [\mu_D^1(x) \quad \mu_D^2(x) \quad \dots \quad \mu_D^c(x)] \quad (15)$$

پس از تعیین این بردار، دسته الگوی x متناظر با ذریه‌ای از بردار $D(x)$ خواهد بود که بیشترین مقدار را دارد. خروجی دسته‌بندها می‌تواند به صورت زیر در ساختار ماتریسی به نام ماتریس نیم‌رخ (پروفایل) تصمیم^۱

¹ decision profile (DP)

که خروجی دسته‌بندها تخمین خوبی از احتمال‌های پسین هر دسته فراهم کنند، این قاعده بهترین تخمین احتمال پسین کلی دسته انتخاب شده را ارائه می‌کند. □ میانگین تعمیم یافته^۲: بسیاری از قواعد بالا درحقیقت نوع خاصی از روش میانگین عمومی زیر هستند:

$$\mu_j(x, \alpha) = \left(\frac{1}{L} \sum_{i=1}^L d_{i,j}(x)^\alpha \right)^{1/\alpha} \quad (22)$$

که به‌ازای مقادیر خاص α قواعد فوق را می‌سازند. برای مثال، اگر

$$\alpha \rightarrow -\infty \Rightarrow \mu_j(x, \alpha) = \min_i \{d_{i,j}(x)\} \quad (23) \text{ قاعده کمینه}$$

$$\alpha = 1 \Rightarrow \mu_j(x, \alpha) = \frac{1}{L} \sum_{i=1}^L d_{i,j}(x) \quad (24) \text{ قاعده میانگین}$$

$$\alpha = 0 \Rightarrow \mu_j(x, \alpha) = \left(\prod_{i=1}^L d_{i,j}(x) \right)^{1/L} \quad (25) \text{ قاعده میانگین هندسی}$$

مثال. مسأله‌ای با ۵ دسته‌بند و ۳ دسته در نظر بگیرید. فرض کنید ماتریس پروفایل زیر برای الگوی x به دست آمده است:

$$DP(x) = \begin{bmatrix} 0.85 & 0.01 & 0.14 \\ 0.30 & 0.50 & 0.20 \\ 0.20 & 0.60 & 0.20 \\ 0.10 & 0.70 & 0.20 \\ 0.10 & 0.10 & 0.80 \end{bmatrix}$$

$$\mu_j(x) = \max_{i=1 \dots L} \{d_{i,j}(x)\} \quad (19)$$

$$\mu_j(x) = \min_{i=1 \dots L} \{d_{i,j}(x)\} \quad (20)$$

$$\mu_j(x) = \text{median}\{d_{i,j}(x)\} \quad (21)$$

در همه این شیوه‌ها تصمیم نهایی، دسته‌ای خواهد بود که بیشترین امتیاز $\mu_j(x)$ را داشته باشد. روش کمینه، محافظه‌کارانه‌ترین روش ترکیب است؛ چرا که دسته‌ای را انتخاب می‌کند که کمترین امتیازی که تمام دسته‌بندها به آن دسته داده‌اند، بیشترین مقدار باشد.

□ میانگین اصلاح شده^۱: زمانی که برخی دسته‌بندها امتیاز نامرسوم کم یا زیادی را به دسته‌ها بدهند، روش ترکیب میانگین ممکن است نتایج خوبی را ایجاد نکند. برای حل این مشکل، بدین‌ترین و خوش‌بین‌ترین دسته‌بندها کنار گذاشته می‌شوند. از این‌رو، در این روش P کمترین امتیاز و بیشترین امتیاز کنار گذاشته می‌شود و شایان توجه است که روش میانگین اصلاح شده با کران 50% معادل با روش میانه است.

□ ضرب: در این روش امتیاز دسته‌بندها در هم ضرب می‌شوند. این روش، به بدبینانه‌ترین دسته‌بند بسیار

دسته‌بندها	0.30	0.25	0.20	0.10	0.15
	دسته‌بند ۱	دسته‌بند ۲	دسته‌بند ۳	دسته‌بند ۴	دسته‌بند ۵
	C ₁ C ₂ C ₃	C ₁ C ₂ C ₃	C ₁ C ₂ C ₃	C ₁ C ₂ C ₃	C ₁ C ₂ C ₃
سطرهای DP(x)	0.85 0.01 0.14	0.3 0.5 0.2	0.2 0.6 0.2	0.1 0.7 0.2	0.1 0.1 0.8
<u>Product Rule:</u>			C ₁ : 0.0179,	C ₂ : 0.00021,	C ₃ : 0.0009
<u>Sum Rule:</u>			C ₁ : 1.55,	C ₂ : 1.91,	C ₃ : 1.54
<u>Mean Rule:</u>			C ₁ : 0.310,	C ₂ : 0.382,	C ₃ : 0.308
<u>Max Rule:</u>			C ₁ : 0.85,	C ₂ : 0.7,	C ₃ : 0.8
<u>Median Rule:</u>			C ₁ : 0.2,	C ₂ : 0.5,	C ₃ : 0.2
<u>Minimum Rule:</u>			C ₁ : 0.1,	C ₂ : 0.01,	C ₃ : 0.14
<u>Weighted Average</u>			C ₁ : 0.395,	C ₂ : 0.333,	C ₃ : 0.272

<u>Majority Voting:</u>			C ₁ : 1,	C ₂ : 3,	C ₃ : 1 Vote
<u>Weighted Majority Voting:</u>			C ₁ : 0.3,	C ₂ : 0.55,	C ₃ : 0.15 Votes
<u>Borda Count:</u>			C ₁ : 5	C ₂ : 6,	C ₃ : 4 Votes

(شکل ۱۱): نتایج ترکیب دسته‌بندها با روش‌های مختلف (برگرفته از (Polikar 2006))

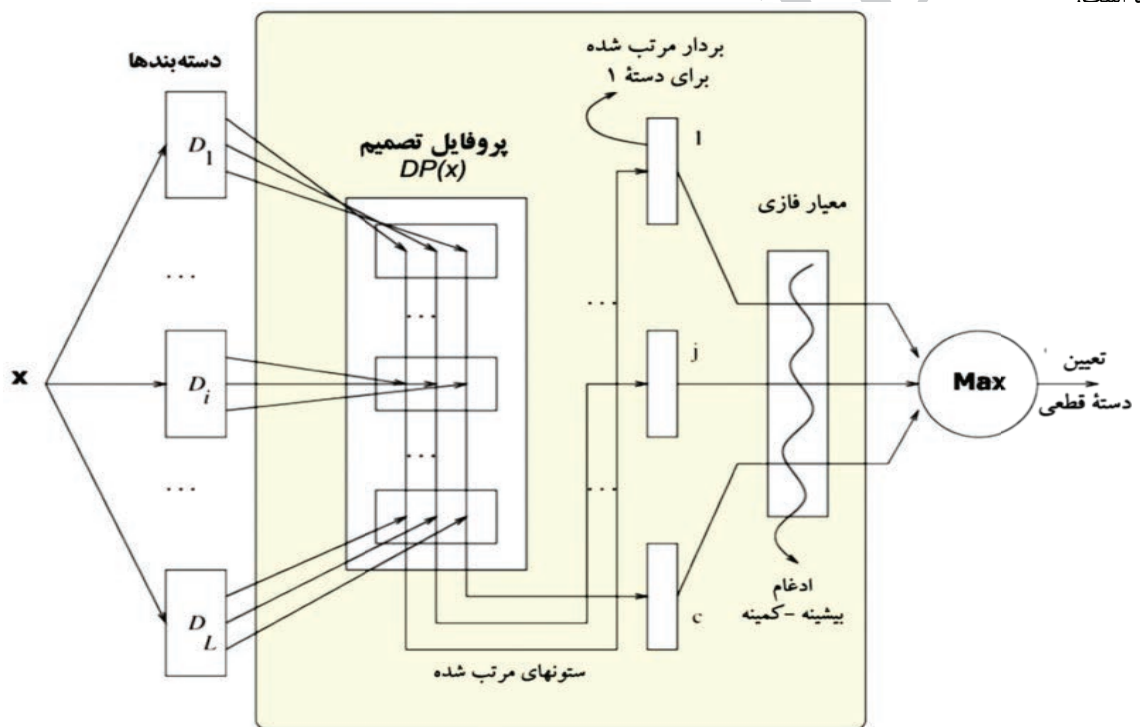
به‌کارگیری روش‌های فوق، نتایج مختلفی را برای عضویت نمونه x به دسته‌های ω_1 ، ω_2 و ω_3 حاصل می‌کند. این نتایج در (شکل ۱۱) نشان داده شده است؛ و دسته

حساس است: یعنی امتیاز کم (نزدیک به صفر) برای یک دسته از سوی هر کدام از دسته‌بندها در عمل می‌تواند احتمال انتخاب آن دسته را از بین ببرد. با این حال، زمانی

² generalized mean

¹ trimmed mean

محاسبه می‌شود که هر بردار متناظر با یک دسته است، و شامل L مقدار از یک معیار فازی است. سپس، L ستون ماتریس پروفایل تصمیم، شامل مقادیر امتیاز L دسته‌بند به دسته w_j ، مرتب می‌شود و با معیار فازی ادغام می‌شود تا میزان تعلق نمونه x به دسته w_j ، $\mu_j(x)$ ، تعیین شود. لذا، ترکیب فازی نوعی جستجوی بیشترین نمره توافقی بین شواهد عینی^۲ (مقادیر خروجی مرتب‌شده دسته‌بندها برای دسته j) و مقادیر موردانتظار (L مقدار معیار فازی) است. توضیحات بیشتر این روش در (سیدحسین نبوی کریزی احسان‌اله کبیری پائیز ۱۳۸۴؛ Kuncheva, 2004) آورده شده است. نمودار بلوکی روش انتگرال فازی در (شکل ۱۲) نشان داده شده است.



(شکل ۱۲): نمودار بلوکی روش انتگرال فازی

۳-۵-۳- الگوهای تصمیم^۳

ایده روش ترکیب الگوهای تصمیم (DT)، به یاد داشتن مرسوم‌ترین پروفایل تصمیم برای هر دسته است که الگوی تصمیم دسته (DT_j) گفته می‌شود؛ و سپس مقایسه آن با پروفایل تصمیم DP با استفاده از یک معیار شباهت است (Kuncheva, Bezdek, and Duin 2001). نزدیک‌ترین شباهت، برچسب نمونه ناشناخته x را تعیین می‌کند. برای

منتخب در هر روش پرننگ نشان داده شده است.

از (شکل ۱۱) چند نکته نتیجه می‌شود:

- ۱) با روش‌های مختلف، دسته‌های مختلف می‌توانند برای الگوی x انتخاب شوند. در این مثال، دسته ۲ در اغلب روش‌ها انتخاب شده است؛ دسته ۱ در سه روش و دسته ۳ تنها در یک روش (روش کمینه) انتخاب شده است.
- ۲) قاعده حاصل ضرب و قاعده کمینه امتیاز دسته ۲ را به شدت می‌کاهند؛ چرا که این دسته امتیاز کمی از دسته‌بند C_1 گرفته است.
- ۳) همان‌گونه که انتظار داشتیم، قاعده میانگین و مجموع نتایج یکسانی را ارائه می‌کنند.
- ۴) قاعده میانگین وزن‌دار دسته ۱ را انتخاب کرده است، چرا که دسته‌بند ۱ که وزن زیادی دارد و امتیاز بالایی به دسته ۱ داده است.

همچنین سه روش ترکیب برای خروجی‌های مطلق در این شکل آورده شده است و هر سه دسته ۲ را انتخاب کرده‌اند (در روش شمارش بوردا، زمانی که وزن دسته‌بند ۱ با دیگر دسته‌بندها برابر است، رتبه‌بندی به نفع دسته ۱ لحاظ شده است).

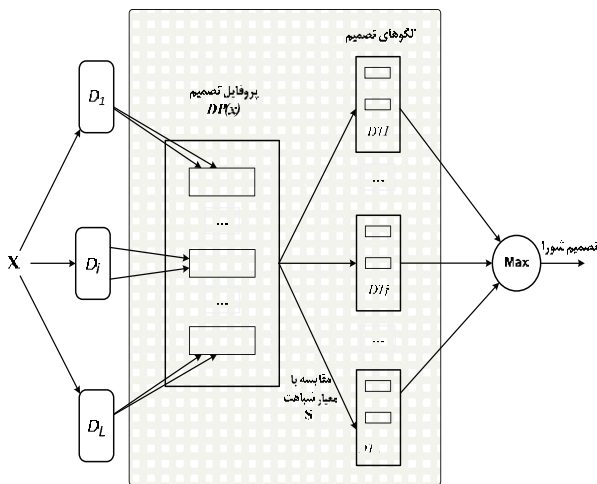
۳-۵-۲- انتگرال فازی^۱

در روش انتگرال فازی (FI) برای نمونه x ، c بردار با طول L

² objective evidence

³ decision templates (DT)

¹ fuzzy integral (FI)



(شکل ۱۴): دیاگرام بلوکی روش الگوهای تصمیم

۳-۵-۴- ترکیب دیمپستر- شفر^۳

این روش از مفهوم هم‌جوشی داده^۴ سرچشمه گرفته است. بسیاری از شیوه‌های هم‌جوشی داده براساس نظریه شواهد دیمپستر- شفر (Dempster 1967; Shafer 1976) هستند که از توابع باور^۵ برای کمی کردن شواهد موجود در منابع اطلاعاتی مختلف استفاده می‌کند و سپس با قاعده ترکیب دیمپستر- شفر ادغام می‌شوند. توضیحات کامل این روش در (نبوی کریزی و کبیر ۱۳۸۴؛ Kuncheva 2004) آورده شده است.

۳-۵-۵- الگوریتم‌های تکاملی

از الگوریتم‌های تکاملی برای ترکیب دسته‌بندها در برخی پژوهش‌ها استفاده شده است. در پژوهش (Lin, Wang, and Yeung 2005) با هدف بهبود شناسایی دست‌خط چینی دو روش ترکیب بر مبنای الگوریتم ژنتیک ارائه شده است. «نبوی کریزی» و همکارانش روش جدیدی بر پایه الگوریتم بهینه‌سازی تجمع ذرات^۶ برای ترکیب چند دسته‌بند شبکه عصبی ارائه کردند (Nabavi-Kerizi, Abadi, and Kabir 2010). نتایج تجربی پژوهش بر روی چند مجموعه داده، توانایی این روش ترکیب را نشان می‌داد. در پژوهشی دیگر (نبوی کریزی و کبیر ۱۳۸۷)، یک روش دومرحله‌ای برای ترکیب دسته‌بندها با ترکیب روش اختلاط خبرگان و PSO ارائه شده است.

³ Dempster-Shafer

^۴ حوزه‌ای از تحلیل داده که اصولاً برای ادغام شواهدی از منابع مختلف

شکل گرفته است.

⁵ belief functions

⁶ particle swarm optimization (PSO)

تعیین شباهت از معیارهای متفاوتی استفاده شده است. دو معیار شباهت مهم عبارتند از:

۱. فاصله مربع اقلیدسی: امتیاز شورا برای دسته ω_j برابر است با:

$$\mu_j(x) = 1 - \frac{1}{L \times C} \sum_{i=1}^L \sum_{k=1}^c [DT_j(i, k) - d_{i,k}(x)]^2 \quad (26)$$

که $DT_j(i, k)$ ، $d_{i,k}(x)$ ، DT_j امتیاز (یا) آمین ذرات الگوی تصمیم $\mu_j(x)$ است. دسته با بیشترین امتیاز (یا) $\mu_j(x)$ تصمیم شورا خواهد بود. شایان ذکر است معیارهای دیگر فاصله مانند فاصله ماهالونبیس^۱ نیز می‌تواند به کار گرفته شود.

۲. تفاوت متقارن^۲: در این معیار که از نظریه فازی منشأ گرفته است (Kuncheva 2001; Kuncheva 2003) امتیاز دسته ω_j برابر است با:

$$\mu_j(x) = 1 - \frac{1}{L \times C} \sum_{i=1}^L \sum_{k=1}^c \max\{\min\{DT_j(i, k), 1 - d_{i,k}(x)\}, \min\{1 - DT_j(i, k), d_{i,k}(x)\}\} \quad (27)$$

(شکل ۱۳) الگوریتم روش الگوهای تصمیم را نشان می‌دهد و (شکل ۱۴) دیاگرام بلوکی روش را نشان می‌دهد. توضیحات بیشتر این روش در (نبوی کریزی و کبیر ۱۳۸۴؛ Kuncheva 2004) آورده شده است.

۱. آموزش

به ازای $j = 1, \dots, c$ میانگین ماتریسهای پروفایل متعلق به دسته ω_j را برای تمام نمونه‌های آموزش (Z) محاسبه کن. ماتریس میانگین هر دسته را الگوی تصمیم DT_j می‌نامیم:

$$DT_j = \frac{1}{N_j} \sum_{z_k \in \omega_j} DP(z_k) \quad (28)$$

که N_j تعداد عناصر (الگوهای) دسته ω_j است.

۲. آزمون

برای داده آزمون $x \in \mathbb{R}^n$ ، ماتریس $DP(x)$ را ایجاد کند. شباهت بین $DP(x)$ و تمام DT_j ها را محاسبه کن؛

$$\mu_j(x) = S(DP(x), DT_j) \quad j = 1, \dots, c \quad (29)$$

(شکل ۱۳): الگوریتم روش ترکیب الگوهای تصمیم

¹ mahalanobis distance

² symmetric difference

۴- اندازه‌گیری گوناگونی

همان‌گونه که بیان شد، کارایی بهتر سیستم‌های شورایی به گوناگونی دسته‌بندهای پایه در شورا وابسته است. به‌گونه‌ای که اگر دسته‌بندهای پایه گوناگون در خطا نباشند، صحت دسته‌بندی شورا بهتر از دسته‌های مجزا نخواهد بود. در مرجع (Brown et al., 2005)، اهمیت گوناگونی دسته‌بندها به‌طور نظری نشان داده شده که با توجه به محدودیت مقاله، به آن پرداخته نشده است. اغلب روش‌های طراحی شورایی دسته‌بندها (که در بخش ۲ معرفی شد)، گوناگونی را به‌طور مستقیم در طراحی روش در نظر نمی‌گیرند؛ بلکه گوناگونی به‌طور ضمنی با استفاده از مجموعه‌های داده یا ویژگی متفاوت حاصل می‌شود. با این حال، برای ارزیابی گوناگونی نهایی شورایی طراحی شده می‌توان از معیارهای گوناگونی مختلف استفاده کرد. در این بخش، معیارهای ارزیابی گوناگونی به‌اجمال معرفی شده است. مطالعه گسترده مفهوم و معیارهای گوناگونی در حوزه دسته‌بندی و رگرسیون در (Brown et al., 2005; Windeatt, 2005) آورده شده است.

معیارهای مختلفی برای ارزیابی کمی گوناگونی تعریف شده است. ساده‌ترین آنها معیارهای زوجی^۱ هستند که بین دو دسته‌بند تعریف می‌شوند. زمانی که تعداد دسته‌بندها L باشد ($L \geq 2$)، تعداد $L(L-1)/2$ معیار جفتی بین دو دسته‌بند محاسبه می‌شود و گوناگونی کل سیستم دسته‌بند چندگانه برابر با میانگین این معیارها خواهد بود. برای دو دسته‌بند h_i و h_j از نمادهای جدول زیر استفاده می‌کنیم:

(جدول ۲): نمادهای دسته‌بندی برای دو دسته‌بند h_i و h_j

	h_j اشتباه است	h_i صحیح است
h_i اشتباه است	d	c
h_i صحیح است	b	a

که a نسبتی از نمونه‌ها است که با هر دو دسته‌بند صحیح دسته‌بندی شده‌اند، b نسبتی از نمونه‌ها است که با دسته‌بند h_i صحیح دسته‌بندی شده‌اند و با دسته‌بند h_j اشتباه دسته‌بندی شده‌اند. نمادهای c و d نیز مطابق جدول به طرز مشابه تعریف می‌شوند. مشخصاً $a+b+c+d=1$ خواهد بود. بر اساس این نمادها معیارهای گوناگونی زیر ارائه شده است:

□ همبستگی^۲: گوناگونی بر اساس همبستگی بین خروجی دو دسته‌بند سنجیده می‌شود که طبق رابطه^۳ ۳۰ تعریف شده است:

$$\rho_{i,j} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}, \quad 0 \leq \rho \leq 1 \quad (30)$$

بیشترین گوناگونی برای $\rho=0$ به دست می‌آید که نشان می‌دهد دو دسته‌بند ناهمبسته‌اند.

□ آماره^۳ Q : این معیار طبق رابطه زیر تعریف می‌شود:

$$Q_{i,j} = \frac{ad - bc}{ad + bc} \quad (31)$$

زمانی که نمونه‌های یکسان با هر دو دسته‌بند صحیح دسته‌بندی شده باشند این آماره مقادیر مثبت و در غیر این صورت مقادیر منفی را اختیار می‌کند. بیشترین گوناگونی برای $Q=0$ به دست می‌آید.

□ معیار آنتروپی: این معیار فرض می‌کند زمانی گوناگونی حداکثر است که برای هر نمونه، نیمی از دسته‌بندها صحیح و نیمی از آنها اشتباه دسته‌بندی کنند. بر اساس این فرض و تعریف ζ_i به‌عنوان تعدادی از دسته‌بندها از میان L دسته‌بند که نمونه x_i را اشتباه دسته‌بندی می‌کنند، معیار آنتروپی به‌صورت زیر تعریف می‌شود:

$$E = \frac{1}{N} \sum_{i=1}^N \frac{1}{L - [L/2]} \min\{\zeta_i, (T - \zeta_i)\} \quad (32)$$

که $[.]$ جزء صحیح عدد است؛ و N بعد مجموعه داده است. معیار آنتروپی بین ۰ و ۱ است؛ که ۰ نشان‌دهنده یکسان بودن تمام دسته‌بندها و ۱ نشان‌دهنده بیشترین گوناگونی است (Kuncheva and Whitaker, 2003).

۵- حوزه‌های پژوهشی فعال

با وجود دو دهه پژوهش، برخی از پژوهشگران معتقدند این حوزه از یادگیری ماشینی، به بلوغ خود رسیده است. با این حال، به نظر می‌رسد سیستم‌های شورایی توجه زیادی را به خود جلب خواهد کرد. این موضوع تا حد زیادی به علت کاربردهای جدیدی است که از سیستم‌های شورایی بهره می‌برند. اخیراً کاربردها و زمینه‌های پژوهشی نوینی در حوزه سیستم‌های دسته‌بند چندگانه شکل گرفته است. در این بخش، برخی از مسیرهای آتی پژوهش و زمینه‌های پژوهشی فعال در حوزه سیستم‌های شورایی که به نظر مؤلفان مقاله فراخور پژوهش بیشتر هستند، آورده شده است:

^۲ correlation^۳ Q-Statistic^۱ pair-wise

۵-۱- یادگیری تدریجی^۱

در برخی مسائل کاربردی، مجموعه داده مسأله در ابتدا موجود نیست و به تدریج در بسته‌های مختلف در دسترس قرار می‌گیرد. علاوه بر این ممکن است نمونه‌های بعدی شامل نمونه‌هایی از دسته‌ای جدید باشد. لذا لازم است بدون فراموش کردن دانش پیشین و بدون نیاز به نمونه‌های قبلی، اطلاعات نمونه‌های جدید نیز یاد گرفته شود. توانایی دسته‌بند برای یادگیری تحت این شرایط، به‌طور معمول یادگیری تدریجی خوانده می‌شود. رویکردی عملی برای یاد گرفتن از نمونه‌های جدید صرف نظر کردن از دسته‌بند موجود و استفاده از دسته‌بندی جدید با استفاده از تمام نمونه‌های است که تاکنون به دست آمده است. با این حال، این رویکرد منجر به از دست دادن تمام دانش به دست آمده تا آن مرحله می‌شود؛ مشکلی که فراموشی فاجعه‌آمیز^۲ خوانده می‌شود. در صورتیکه که زمان یا هزینه محاسباتی زیادی برای آموزش دوباره دسته‌بند لازم باشد، این رویکرد مطلوب نیست. مهم‌تر از این، اگر مجموعه داده اولیه حذف شده یا آسیب دیده باشد یا به‌طور کلی در دسترس نباشد^۳ این رویکرد عملی نیست.

استفاده از سیستم‌های شورایی در این مسائل منجر به نتایج خوبی شده است. ایده اصلی، ایجاد شوراهایی جدید از دسته‌بندها با هر مجموعه داده جدید و سپس ترکیب خروجی آنها با استفاده از یکی از روش‌های معمول است. الگوریتم Learn⁺⁺ مهم‌ترین الگوریتم ارائه شده در این زمینه است که در کاربردهای مختلف، حتی زمانی که نمونه‌های بعدی شامل نمونه‌هایی از دسته‌هایی جدید باشد، نتایج خوبی را نشان داده است (Polikar et al. 2004; R. Polikar et al. 2001).

۵-۲- توسعه سیستم‌های شورایی در انواع

دیگر مدل‌های یادگیری

اغلب پژوهش‌ها در سیستم‌های شورایی در حوزه دسته‌بندی ساده انجام شده است. مطالعات بیشتری برای به کارگیری این رویکرد در حوزه‌های دیگر یادگیری ماشین مانند خوشه‌بندی، رگرسیون، سری زمانی و دیگر حوزه‌های یادگیری با نظارت مانند یادگیری فعال^۴ یا دسته‌بندی توالی^۵ زمینه‌های پژوهشی مناسبی خواهد بود.

¹ incremental learning

² catastrophic forgetting

³ این حالتها در بسیاری از مسائل پزشکی یا نظامی که دسترسی به مجموعه داده محدود یا محرمانه است، مرسوم است.

⁴ active learning

⁵ sequence classification

۵-۳- ارتباط گوناگونی با مسائل دسته‌بندی

روابط ریاضی بین کارایی سیستم شورا با گوناگونی آن برای مسائل رگرسیون ارائه شده است (Brown et al. 2005). با این حال، تاکنون این گونه روابط برای مسائل دسته‌بندی به دست نیامده است.

۵-۴- انتخاب دسته‌بند

رویکرد مشابه ترکیب دسته‌بندها در سیستم‌های شورایی، انتخاب دسته‌بند^۶ است (Woods, Kegelmeyer, and Bowyer 1997). در ادغام دسته‌بندها، هر دسته‌بند از تمام نمونه‌های فضای ویژگی استفاده می‌کند. در این حالت، فرایند ادغام شامل ترکیب دسته‌بندهای پایه و ضعیف‌تر برای دست‌یابی به دسته‌بند با کارایی بیشتر است. در انتخاب دسته‌بند، هر دسته‌بند در شورا بخشی از فضای ویژگی را به خوبی یاد می‌گیرد و لذا دسته‌بندی نمونه‌های آن بخش از فضا را انجام می‌دهد. از این رو، در ادغام دسته‌بندها از روش‌های ترکیب (که در فصل پیش معرفی شده است) برای تعیین دسته الگوی x استفاده می‌کنیم؛ درحالی‌که در رویکرد انتخاب، به‌طور معمول یک دسته‌بند انتخاب می‌شود (Woods, Kegelmeyer, and Bowyer, 1997).

رویکرد انتخاب دسته‌بند همانند ادغام دسته‌بند توجه زیادی را به خود جلب نکرده است. این مسأله ممکن است در آینده تغییر کند؛ چرا که روش انتخاب دسته‌بند کارایی خوبی را نشان داده است (Kuncheva, 2004). دسته‌بندهای آبشاری^۷ (Zhang, Bui, and Suen, 2007) نیز از دیگر روش‌های رویکرد انتخاب ویژگی است که با وجود توانایی آنها در کاربردهای عملی، پژوهش‌های بسیار کمی به آن پرداخته‌اند. در این روش، تنها یک دسته‌بند در هر لحظه فعال است و سایر دسته‌بندها غیرفعال هستند. برای دسته‌بندی نمونه x ، اولین دسته‌بند سعی در دسته‌بندی آن می‌کند. اگر دسته‌بند در تصمیم خود، مطمئن^۸ است، x برچسب‌گذاری می‌شود و فرآیند متوقف می‌شود. در غیر این صورت، نمونه x به دسته‌بند بعدی می‌رود و این رویه ادامه می‌یابد.

۵-۵- تخمین اطمینان^۹

استفاده از سیستم‌های شورایی همچنین می‌تواند برای ارزیابی کمی میزان اطمینان تصمیم به کار گرفته شود. به

⁶ classifier selection

⁷ cascade classifier

⁸ certain

⁹ confidence estimation

- R. P.W. Duin, D. Ridder, P. Juszczak, P. Paclik, E. Pekalska and D. Tax, "PRTools," 2005. Available online at: <http://www.prtools.org>
- D. Stork and E. Yom-Tov, "Computer Manual in Matlab to Accompany Pattern Classification, 2nd Edition, New York: Wiley, 2004.
- همچنین در دسترس بودن مهم‌ترین الگوریتم‌های شورایی در نرم‌افزارهای معروف متن‌باز در (جدول ۳) نشان داده شده است:

(جدول ۳): در دسترس بودن الگوریتم‌های مهم ایجاد شورا در نرم‌افزارهای متن‌باز

الگوریتم	نرم‌افزارهای موجود
AdaBoost	Weka, Orange, Tanagra, RapidMiner, R
Bagging	All
Random Forest	All
Wagging	Weka, RapidMiner
Attribute Bagging	Weka
Stacking	Weka, Tanagra, RapidMiner
ECOC	Weka, RapidMiner

شایان ذکر است برای استفاده سایر پژوهش‌گران در این حوزه، گُذ پیاده‌سازی ده روش مهم ترکیب شورا نوشته شده که از پایگاه <http://www.mathworks.com/matlabcentral/fileexchange/> قابل دسترسی است.

۷- منابع

Allwein, Erin L., Robert E. Schapire, and Yoram Singer. 2001. Reducing multiclass to binary: a unifying approach for margin classifiers. *Journal of Machine Learning Research* 1:113-141.

Bagheri, Mohammad Ali, and Gholam Ali Montazer. 2011. Ensemble Classifier Strategy Based on Transient Feature Fusion in Electronic Nose. Paper read at 14th international symposium on olfaction and electronic nose, at New York City, NY, USA.

Banfield, R. E., L. O. Hall, K. W. Bowyer, and W. P. Kegelmeyer. 2007. A Comparison of Decision Tree Ensemble Creation Techniques. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29 (1):173-180.

سال ۱۳۹۰ شماره ۲ پیاپی ۱۶

زبان ساده، اگر اغلب دسته‌بندها در دسته‌بندی یک الگو تصمیم یکسانی داشته باشند، این تصمیم شورا اطمینان بالایی خواهد داشت. متقابلاً، اگر تصمیم نهایی براساس تفاوت زیاد نظرات دسته‌بندها گرفته شده باشد، تصمیم شورا اطمینان کمی خواهد داشت. نشان داده شده که خروجی‌های امتیازدار دسته‌بندها می‌تواند برای تخمین احتمال پسین تصمیم استفاده شود؛ که به نوعی نشان‌دهنده اطمینان تصمیم است (Muhlbaier, Topalis, and Polikar 2005).

۶- جمع‌بندی و نتیجه‌گیری

در این مقاله تلاش شده است سیستم‌های دسته‌بند چندگانه (سیستم‌های شورایی) به‌عنوان یکی از زمینه‌های فعال در حوزه یادگیری ماشین معرفی شود. دو بخش اساسی در این سیستم‌ها که دو عامل تعیین‌کننده موفقیت آنهاست به‌اجمال بیان شده است. عامل مهم اول، ایجاد گوناگونی در شورا است که با روش‌های مختلف طراحی شورای دسته‌بندها ایجاد می‌شود. رویکردهای زیرنمونه، زیرفضا و رویکرد ایجاد شورا با دسته‌بندهایی از انواع مختلف معرفی شده است. عامل مهم دوم، روش ترکیب دسته‌بندها است: روش‌های مختلف ترکیب براساس نوع خروجی دسته‌بندها تقسیم‌بندی شده و به‌اختصار آورده شده است. همچنین سعی شده هر روش با ذکر مثالی ساده توضیح داده شود. این دو عامل مهم به‌گونه‌ای مکمل یکدیگرند. به‌گونه‌ای که مجموعه ایده‌آلی از دسته‌بندها می‌تواند تا حد زیادی ایجاد قاعده ترکیب را ساده کند و قاعده ترکیبی قوی می‌تواند حتی با شورای دسته‌بندهای ضعیف، نتایج خوبی را بدهد. در پایان با بررسی پژوهش‌های اخیر در این حوزه، زمینه‌های پژوهشی پیشنهادی ارائه شده است. امید است این مقاله چراغ راهی برای شروع پژوهش‌های آتی در حوزه سیستم‌های شورایی باشد.

پیوست الف) نرم‌افزارهای پیاده‌سازی الگوریتم‌های شورایی

بسته‌های نرم‌افزاری زیر شامل توابع بر مبنای MATLAB هستند که می‌توان برای پیاده‌سازی الگوریتم‌های شورا استفاده شوند:

- C. Merkwinh and J. Wichard, "ENTOO—A Matlab Toolbox for Ensemble Modeling," 2002. Available online at: <http://chopin.zet.agh.edu.pl/~wichtel>

induced by multivalued mappings. *Annals of Mathematical Statistics* 38 (2):325–339.

Dietterich, T. G. 1997. Machine learning research: Four current directions. *Artificial Intell. Mag.* 18(4):97-136.

Dietterich, T.G., and G. Bakiri. 1995. Solving Multiclass Learning Problems via Error-Correcting Output Codes. *Journal of Artificial Intelligence Research* 2:263–286.

Dietterich, Thomas. 2000. Ensemble Methods in Machine Learning. *Lecture Notes in Computer Science* 1857:1-15.

Ebrahimpour, Reza, Ehsanollah Kabir, Hossein Esteky, and Mohammad Reza Yousefi. 2008. View-independent face recognition with Mixture of Experts. *Neurocomputing* 71 (4-6):1103-1107.

Ebrahimpour, Reza, Ehsanollah Kabir, and Mohammad Yousefi. 2011. Improving mixture of experts for view-independent face recognition using teacher-directed learning. *Machine Vision and Applications* 22 (2):421-432.

Finlay, Steven. 2011. Multiple classifier architectures and their application to credit risk assessment. *European Journal of Operational Research* 210 (2):368-378.

Freund, Y., and R. Schapire. 1996. Experiments with a new boosting algorithm. Paper read at Proceeding of the Thirteenth International Conference on Machine Learning.

Freund, Y., and R.E. Schapire. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* 55 (1):119–139.

Fumera, G., F. Roli, and A. Serrau. 2008. A Theoretical Analysis of Bagging as a Linear Combination of Classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 30 (7):1293-1299.

Gey, Servane, and Jean-Michel Poggi. 2006. Boosting and instability for regression trees. *Computational Statistics & Data Analysis* 50 (2):533-550.

Giacinto, G., and F. Roli. 2001. An approach to the automatic design of multiple classifier systems. *Pattern Recognition Letters* 22:25–33.

Bao, Y., and N. Ishii. 2002. Combining multiple K-nearest neighbor classifiers for text classification by reducts. Paper read at Proceedings of the 5th International Conference on Discovery Science, *Lecture Notes in Computer Science*, at Berlin.

Bauer, Eric, and Ron Kohavi. 1999. An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants. *Machine Learning* 36 (1):105-139.

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

Breiman, L. 1996. Bagging Predictors. *Machine Learning* 24:123-140.

———2001. Random Forests. *Machine Learning* 45 (1): 5–32.

Brodley, arla, and Terran Lane. 1996. Creating and Exploiting Coverage and Diversity. Paper read at Proceedings of AAAI-96 Workshop on Integrating Multiple Learned Models, , at Portland, OR.

Brown, Gavin, Jeremy Wyatt, Rachel Harris, and Xin Yao. 2005. Diversity creation methods: a survey and categorisation. *Information Fusion* 6 (1):5-20.

Bryll, R., R. Gutierrez-Osuna, and F. Quek. 2003. Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets. *Pattern Recognition* 36 1291–1302.

———2003. Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets. *Pattern Recognition* 36 1291–1302.

Chen, Shih-Chieh, Shih-Wei Lin, and Shuo-Yan Chou. 2011. Enhancing the classification accuracy by scatter-search-based ensemble approach. *Applied Soft Computing* 11 (1):1021-1028.

Cunningham, P., and J. Carney. 2000. Diversity versus quality in classification ensembles based on feature selectio. In R.L. de Mántaras, E. Plaza (Eds.), *Proceedings of the ECML 2000*. Barcelona, Spain: Springer, Berlin.

Dasarathy, B. V., and B. V. Sheela. 1979. A composite classifier system design: Concepts and methodology. *Proceedings of the IEEE* 67 (5):708-713.

Dempster, A.P. 1967. Upper and lower probabilities

computation 6(2)-181-214.

Kim, Myoung-Jong, and Dae-Ki Kang. 2010. Ensemble with neural networks for bankruptcy prediction. *Expert Systems with Applications* 37 (4):3373-3379.

Kotsiantis, S., K. Patriarcheas, and M. Xenos. 2010. A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education. *Knowledge-Based Systems* 23 (6):529-535.

Krogh, A., and J. Vedelsby. 1995. Neural network ensembles, cross validation, and active learning. In *Advances in Neural Information Processing Systems*, edited by D. Touretzky and T. Leen. Cambridge, MA: MIT Press.

Kuncheva, L. I. 2003. "Fuzzy" versus "nonfuzzy" in combining classifiers designed by Boosting. *Fuzzy Systems, IEEE Transactions on* 11 (6):729-741.

———. 2004. *Combining Pattern Classifiers: Methods and Algorithms*. New York, NY: Wiley.

Kuncheva, L. I., J. C. Bezdek, and R. P. W. Duin. 2001. Decision Templates for Multiple Classifier Fusion: An Experimental Comparison. *Pattern Recognition* 34 (2):299-314.

Kuncheva, L. I., and L. C. Jain. 2000. Designing classifier fusion systems by genetic algorithms. *Evolutionary Computation, IEEE Transactions on* 4 (4):327-336.

Kuncheva, L. I., M. Skurichina, and R. P. W. Duin. 2000. An experimental study on diversity for bagging and boosting with linear classifiers. *Information Fusion* 3 (4):245-258.

Kuncheva, L., and C. Whitaker. 2003. "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning* 51 (2):181-207.

Kuncheva, Ludmila I. 2001. Using measures of similarity and inclusion for multiple classifier fusion by decision templates. *Fuzzy Sets and Systems* 122 (3):401-407.

Lam, L., and C. Y. Suen. 1997. Application of majority voting to pattern recognition: An analysis of its behavior and performance. *IEEE Transactions on Systems, Man, and Cybernetics* 27 (5):553-568.

Govindarajan, M., and R. M. Chandrasekaran. Intrusion detection using neural based hybrid classification methods. *Computer Networks In Press*, Corrected Proof.

Hansen, L. K., and P. Salamon. 1990. Neural network ensembles. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 12 (10):993-1001.
Hansen, L., and P. Salamon. Oct. 1990. Neural network ensembles. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 12 (10):993-1001.

Hashem, S. 1997. Optimal linear combinations of neural networks: an overview., *Neural Networks* 10 (4):599-614.

Haykin, S., and N. Network. 2004. *A comprehensive foundation. Neural Networks* 2.

Ho, Tin. K. 1998. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20:832-844.

Hong, Yi, Sam Kwong, Yuchou Chang, and Qingsheng Ren. 2008. Unsupervised feature selection using clustering ensembles and population based incremental learning algorithm. *Pattern Recognition* 41 (9):2742-2756.

Hu, Q.H., D.R. Yu, and M.Y. Wang. 2005. Constructing rough decision forests. Paper read at D. Slezak et al. (Ed.), *RSFDGrC 2005, Lecture Notes in Artificial Intelligence*, at Berlin.

Hu, Qinghua, Daren Yu, Zongxia Xie, and Xiaodong Li. 2007. EROS: Ensemble rough subspaces. *Pattern Recognition* 40 3728 – 3739.

Huang, Y.S., and C.Y. Suen. 1993. Behavior-knowledge space method for combination of multiple classifiers. Paper read at *IEEE Computer Vision and Pattern Recog.*

———. 1995. A method of combining multiple experts for the recognition of unconstrained handwritten numerals. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17 (1):90-94.

Jacobs, R.A., M.I. Jordan, S.J. Nowlan, and G.E. Hinton. 1991. Adaptive mixtures of local experts. *Neural Computation* 3:79-87.

Jordan, M.I., and R.A. Jacobs. 1994. Hierarchical mixtures of experts and the EM algorithm. *Neural*

signals. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control* 51 (8):990–1001.

R. Polikar, L. Udpa, S.S. Udpa, and V. Honavar. 2001. Learn++: An incremental learning algorithm for supervised neural networks. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews* 31 (4):497–508.

Rodriguez, J.J., L.I. Kuncheva, and C.J. Alonso. 2006. Rotation forest: A new classifier ensemble method. *IEEE Trans. Pattern Anal. Machine Intell* 28 (10):1619–1630.

Rokach, Lior. 2008. Genetic algorithm-based feature set partitioning for classification problems. *Pattern Recognition* 41 (5):1676 - 1700.

———. 2008. Mining manufacturing data using genetic algorithm/C45-based feature set decomposition. *Int. J. Intell. Syst. Technol. Appl.* 4 (1/2):57-78.

Ruta, D., and B. Gabrys. 2000. An Overview of Classifier Fusion Methods. *Computing and Information Systems* 7 (1):1-10.

Schapire, Robert E. 1990. The strength of weak learnability. *Machine Learning* 5 (2):197-227.

Shafer, G. 1976. *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton Univ. Press.

Sharkey, Amanda, Noel Sharkey, Uwe Gerecke, and G. Chandroth. 2000. The “Test and Select” Approach to Ensemble Combination. In *Multiple Classifier Systems*: Springer Berlin / Heidelberg.

Shi, Lei, Xinming Ma, Lei Xi, Qiguo Duan, and Jingying Zhao. 2011. Rough set and ensemble learning based semi-supervised algorithm for text classification. *Expert Systems with Applications* 38 (5):6300-6306.

Skurichina, M., and Robert P. W. Duin. 2002. Bagging, Boosting and the Random Subspace Method for Linear Classifiers. *Pattern Analysis & Applications* 5:121-135.

Tsymbal, A., and S. Puuronen. 2002. Ensemble feature selection with the simple Bayesian classification in medical diagnostics. Paper read at Proceedings of the 15th IEEE Symposium on Computer-Based Medical Systems CBMS'2002, at

Lin, Lei, Xiaolong Wang, and Daniel Yeung. 2005. Combining multiple classifiers based on a statistical method for handwritten chinese character recognition. *International Journal of Pattern Recognition and Artificial Intelligence* 19 (8):1027-1040.

Meyer, David, Friedrich Leisch, and Kurt Hornik. 2003. The support vector machine under test. *Neurocomputing* 55 (1-2):169-186.

Muhlbaier, M., A. Topalis, and R. Polikar. 2005. Ensemble confidence estimates posterior probability. In 6th Int. Workshop on Multiple Classifier Sys., edited by N. Oza, R. Polikar, J. Kittler and F. F. Roli.

Nabavi-Kerizi, S. H., M. Abadi, and E. Kabir. 2010. A PSO-based weighting method for linear combination of neural networks. *Computers & Electrical Engineering* 36 (5):886-894.

Nanni, Loris, and Alessandra Lumini. 2009. Ensemble generation and feature selection for the identification of students with learning disabilities. *Expert Systems with Applications* 36 (2, Part 2):3896-3900.

Opitz, David W. 1999. Feature selection for ensembles. In Proc. 16th National Conf. on Artificial intelligence, AAAI. Orlando, Florida, United States: American Association for Artificial Intelligence.

Oza, N.C., and K. Tumer. 2001. Input decimation ensembles: Decorrelation through dimensionality reduction. In 2nd Int. Workshop on Multiple Classifier Systems in Lecture Notes in Computer Science, edited by J. Kittler and F. Roli.

Peng, Yi, Guoxun Wang, Gang Kou, and Yong Shi. 2001. An empirical study of classification algorithm evaluation for financial risk prediction. *Applied Soft Computing* 11 (2):2906-2915.

Pino-Mejías, Rafael, María Dolores Cubiles-de-la-Vega, María Anaya-Romero, Antonio Pascual-Acosta, Antonio Jordán-López, and Nicolás Bellinfante-Crocci. 2010. Predicting the potential habitat of oaks with data mining models and the R system. *Environmental Modelling & Software* 25 (7):826-836.

Polikar, R. 2006. Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine* 6 (3):21-45.

Polikar, R., L. Udpa, S. Udpa, and V. Honavar. 2004. An incremental learning algorithm with confidence estimation for automated identification of NDE

of combining multiple classifiers and their application to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics* 22:418–435.

Yang, L.Y., Z. Qin, and R. Huang. 2004. Design of a multiple classifier system. Paper read at IEEE Proceedings of the Third International conference on Machine Learning and Cybernetics, August 26–29, at Shanghai.

Yu, Enzhe, and Sungzoon Cho. 2006. Ensemble based on GA wrapper feature selection. *Computers & Industrial Engineering* 51 (1):111-116.

Zhang, Chun-Xia, and Jiang-She Zhang. 2008. RotBoost: A technique for combining Rotation Forest and AdaBoost. *Pattern Recognition Letters* 29(10):1524-1536.

Zhang, Ping, Tien D. Bui, and ChingY. Suen. 2007. A novel cascade ensemble classifier system with a high recognition performance on handwritten digits. *Pattern Recognition* 40 3415 – 3429.

نبوی کریزی، سید حسن، و احسان اله کبیر، ۱۳۸۴، ترکیب طبقه‌بندها: ایجاد گوناگونی و قواعد ترکیب. علوم و مهندسی کامپیوتر مجلد ۳، شماره ۳ (الف)، صفحات ۹۵–۱۰۷
 نبوی کریزی، سید حسن، و احسان اله کبیر، ۱۳۸۷. یک روش دو مرحله ای برای ترکیب طبقه‌بندها. نشریه مهندسی برق و مهندسی کامپیوتر ایران ۶ (۱):۶۳–۷۰

Maribor, Slovenia.

Tsymbal, Alexey, Mykola Pechenizkiy, and Pádraig Cunningham. 2005. Diversity in search strategies for ensemble feature selection. *Information Fusion* 6 (1):83-98.

Tumer, K., and J. Ghosh. August 1996. Classifier combining: analytical results and implications, . Paper read at 13th National Conference on Artificial Intelligence, at Portland, Oregon.

Tumer, K., and N.C. Oza. 2003. Input decimated ensembles. *Pattern Anal. Appl.* 6:65–77.

Tutz, Gerhard, and Harald Binder. 2007. Boosting ridge regression. *Computational Statistics & Data Analysis* 51 (12):6044-6059.

Wanas, N. M., and M. S. Kamel. 2001. Decision fusion in neural network ensembles. Paper read at Proceedings of international joint conference on neural networks.

Wang, Shi-jin, Avin Mathew, Yan Chen, Li-feng Xi, Lin Ma, and Jay Lee. 2009. Empirical analysis of support vector machine ensemble classifiers. *Expert Systems with Applications* 36 (3, Part 2):6466-6476.

Wanga, Shu-Lin, Xueling Li, Shanwen Zhang, Jie Gui, and De-Shuang Huang. 10. Tumor classification by combining PNN classifier ensemble with neighborhood rough set based gene reduction. *Computers in Biology and Medicine* 40:179–189.

Windeatt, Terry. 2005. Diversity measures for multiple classifier system analysis and design. *Information Fusion* 6 (1):21-36.

Windeatt, Terry, and Reza Ghaderi. 2003. Coding and decoding strategies for multi-class learning problems. *Information Fusion* 4 (1):11-21.

Woods, K., W.P.J. Kegelmeyer, and K. Bowyer. 1997. Combination of multiple classifiers using local accuracy estimates. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19 (4):405–410.

Wu, QingXiang, and David Bell. 2003. Multi-knowledge Extraction and Application. In *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, edited by G. Wang, Q. Liu, Y. Yao and A. Skowron: Springer Berlin / Heidelberg.

Xu, L., A. Krzyzak, and C. Y. Suen. 1992. Methods



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 ترتیب در سال‌های ۱۳۶۴ و ۱۳۶۹ دریافت کرد.

وی اکنون استاد دانشکده مهندسی برق و کامپیوتر دانشگاه تربیت مدرس است. زمینه پژوهشی مورد علاقه ایشان بازشناسی الگو، به ویژه بازشناسی متون چاپی و دست‌نویس است.

نشانی رایانامک ایشان عبارتست از:

kabir@modares.ac.ir



محمدعلی باقری در سال ۱۳۸۴

مدرک کارشناسی خود را در رشته مهندسی صنایع دانشگاه صنعتی شریف و در سال ۱۳۸۷ مدرک کارشناسی ارشد خود را در رشته مهندسی فناوری اطلاعات از دانشگاه

تربیت مدرس دریافت نمود. وی در حال حاضر دانشجوی دکترای مهندسی فناوری اطلاعات در دانشگاه تربیت مدرس است. زمینه‌های پژوهشی مورد علاقه ایشان شامل شناسایی الگو به‌ویژه سیستم‌های شورایی، بویایی هوشمند (بینی الکترونیکی) و روش‌های بهینه‌سازی تکاملی است.

نشانی رایانامک ایشان عبارتست از:

ma.bagheri@gmail.com

غلام‌علی منتظر در سال ۱۳۴۸ در



کازرون (فارس) به دنیا آمد. وی در سال ۱۳۷۰ مدرک کارشناسی خود را در رشته مهندسی برق از دانشگاه صنعتی خواجه نصیرالدین طوسی و سپس در سال ۱۳۷۳ و ۱۳۷۷ مدارک کارشناسی

ارشد و دکتری خود را در همین رشته از دانشگاه تربیت مدرس اخذ کرد. پس از اتمام تحصیل به عضویت هیئت علمی دانشگاه تربیت مدرس درآمد و در حال حاضر دانشیار مهندسی فناوری اطلاعات در این دانشگاه است. حوزه‌های تخصصی ایشان شامل نرم‌رایانش (نظریه مجموعه‌های فازی، شبکه‌های عصبی مصنوعی، نظریه مجموعه‌های نادقیق) و کاربرد آن در سیستم‌های اطلاعاتی (همچون سیستم یادگیری الکترونیکی و تجارت الکترونیکی) است. وی تاکنون بیش از ۷۰ مقاله در نشریات معتبر علمی و بیش از ۱۵۰ مقاله در کنفرانس‌های معتبر علمی ملی و بین‌المللی منتشر کرده است. علاوه بر این حائز دریافت جوایزی معتبر علمی از جمله «برگزیده جشنواره بین‌المللی خوارزمی»، «برنده کتاب سال دانشگاهی ایران»، «پژوهشگر برگزیده آیسسکو» و «متخصص برجسته فناوری اطلاعات ایران» شده است.

نشانی رایانامک ایشان عبارتست از:

montazer@modars.acir

Archive