یک روش دومرحله‌ای برای ترکیب طبقه‌بندها

سید حسن نبوی کریزی و احسان‌اله کبیر

چکیده: یادگیری دسته‌جمعی، یک رویکرد مؤثر در یادگیری ماشینی است که در آن با ترکیب نتایج چند طبقه‌بند سعی می‌شود تقریب بهتری از یک طبقه‌بند بهینه فراهم شود. برای آنکه ترکیب نتایج طبقه‌بندها مفید واقع شود باید طبقه‌بندهای پایه ضمن برخورداری از کارآیی قابل قبول، دارای خطاهای متفاوتی بوده و قاعده مناسبی برای ترکیب نتایج آنها به کار گرفته شود. در این مقاله یک روش دومرحله‌ای برای ترکیب نتایج طبقه‌بندها پیشنهاد می‌شود که در مرحله اول آن، با روش اختلاط خبره‌ها یک مجموعه طبقه‌بند با خطاهای متفاوت ایجاد می‌شود و در مرحله دوم با استفاده از الگوریتم بهینه‌سازی گروه ذرات، وزن‌های بهینه برای ترکیب خطی نظرات آنها پیدا می‌شوند. نتایج آزمایش‌های ما بر روی چند مجموعه داده متداول، نشان می‌دهند که روش پیشنهادی ما باعث افزایش کارآیی سیستم طبقه‌بندی مرکب نسبت به روش‌های یادگیری مستقل و روش اختلاط خبره‌ها می‌شود.

کلید واژه: **ایجاد گوناگونی، اختلاط خبره‌ها، گروه ذرات، ترکیب طبقه‌بندها، ترکیب خطی، بهینه‌سازی.**

۱- مقدمه

طبقه‌بندی فرآیندی است که در آن هر الگوی ناشناخته بر اساس ویژگی‌های آن، به یکی از کلاس‌های شناخته‌شده نسبت داده می‌شود. به عبارت دیگر، طبقه‌بندی یک نگاشت از فضای n بعدی ویژگی‌ها به فضای K بعدی برچسب‌های کلاسی است که در آن میزان تعلق بردار ویژگی $\mathbf{x} \in \Re^n$ به کلاس‌های مختلف، به صورت یک مقدار عددی بیان می‌شود. طبقه‌بند معمولاً در یک فرآیند یادگیری ساخته می‌شود. بسیاری از الگوریتم‌های یادگیری در حقیقت یک نوع جستجوی محلی انجام می‌دهند که ممکن است در کمینه‌های محلی گرفتار شوند. در صورت گرفتارشدن در کمینه‌های محلی، نمی‌توان طبقه‌بند بهینه داشت. اجرای الگوریتم‌های یادگیری تحت شرایط متفاوت، یادگیری دسته‌جمعی[1]، روشی است برای آنکه بتوان تقریب بهتری را از یک طبقه‌بند بهینه فراهم کرد. در یادگیری دسته‌جمعی، هر الگوریتم یادگیری با توجه به مقدار پارامترهایش، به پاسخ متفاوتی برای مسأله می‌رسد و انتظار می‌رود با ترکیب این پاسخ‌ها، دقت طبقه‌بندی افزایش پیدا کند [۱]. به همین علت، در سال‌های اخیر استفاده از نتایج چند طبقه‌بند، به عنوان یک رویکرد مؤثر در بازشناسی الگو، توجه محققین زیادی را به خود جلب کرده است [۲] و از آن در شاخه‌های مختلف علوم استفاده شده است. تشخیص نفوذ در شبکه‌های کامپیوتری [۳]، شناسایی کدپستی [۴]، بازشناسی دستنوشته [۵]، تشخیص هویت [۶] و شناسایی گوینده [۷] نمونه‌هایی از کاربرد ترکیب طبقه‌بندها هستند.

به طبقه‌بندهایی که نتایج آنها با هم ترکیب می‌شوند طبقه‌بندهای پایه[2] و به مجموعه طبقه‌بندها، سیستم مرکب گفته می‌شود. شبکه‌های عصبی از متداول‌ترین طبقه‌بندهای پایه هستند [۸].

نتایج تئوری [۹] و [۱۰] و تجربی [۱۱] تا [۱۳] نشان می‌دهند که زمانی یادگیری دسته‌جمعی از یادگیری بهترین طبقه‌بند پایه بهتر است که طبقه‌بندهای پایه ضمن برخورداری از کارآیی قابل قبول، گوناگون[3] در خطا بوده و قاعده مناسبی برای ترکیب نتایج آنها به کار گرفته شود. دو طبقه‌بند، زمانی گوناگونی در خطا دارند که الگوهایی که به صورت نادرست طبقه‌بندی می‌کنند متفاوت باشند. تفاوت در موارد خطای طبقه‌بندهای پایه، باعث می‌شود که طبقه‌بندها خطاهای یکدیگر را بپوشانند. به همین علت گوناگونی در خطا، از نکات اساسی در موفقیت سیستم‌های طبقه‌بندی مرکب است [۱۴]. برای ایجاد گوناگونی در خطا، روش‌های متعددی پیشنهاد شده است. این روش‌ها به دو دسته ضمنی[4] و صریح[5] تقسیم می‌شوند.

روش‌های ضمنی، با تغییرات ضمنی در فرآیند یادگیری طبقه‌بندهای پایه، سعی در گوناگون‌کردن آنها دارند. استفاده از مجموعه‌های یادگیری متفاوت [۱۵]، ساختارهای مختلف [۱۶]، بازنمایی‌های متفاوت برای بیان الگو [۱۷]، افزودن نویز به داده‌های ورودی [۱۸] و انجام تبدیلات غیرخطی بر روی داده‌های ورودی [۱۹] از جمله روش‌های ضمنی هستند. متداول‌ترین این روش‌ها، انتخاب تصادفی همراه با جایگزینی نمونه‌ها از بین کلیه نمونه‌های آموزشی است که روش کیسه‌کردن[6] نامیده می‌شود.

روش‌های صریح، با تحت تأثیر قرار دادن مسیر یادگیری طبقه‌بندهای پایه، آنها را با یکدیگر گوناگون می‌سازند. این روش‌ها در فرآیند یادگیری طبقه‌بندها، معیاری از گوناگونی را اعمال کرده و بر اساس آن مسیر یادگیری طبقه‌بندها را در فضای یادگیری طوری تغییر می‌دهند که طبقه‌بندها گوناگون شوند. روش‌های تقویتی[7] [۲۰] و روش‌های جریمه‌ای[8] [۱۴] از مهم‌ترین روش‌های صریح برای ایجاد گوناگونی در طبقه‌بندهای پایه هستند. در روش‌های تقویتی، توزیع داده‌های آموزشی هر طبقه‌بند بر اساس خطاهای طبقه‌بند مرحله قبل تعیین می‌شود [۲۱]. روش‌های جریمه‌ای، با به کارگیری یک مؤلفه جریمه در تابع خطای طبقه‌بندهای پایه، به ایجاد گوناگونی در آنها می‌پردازند.

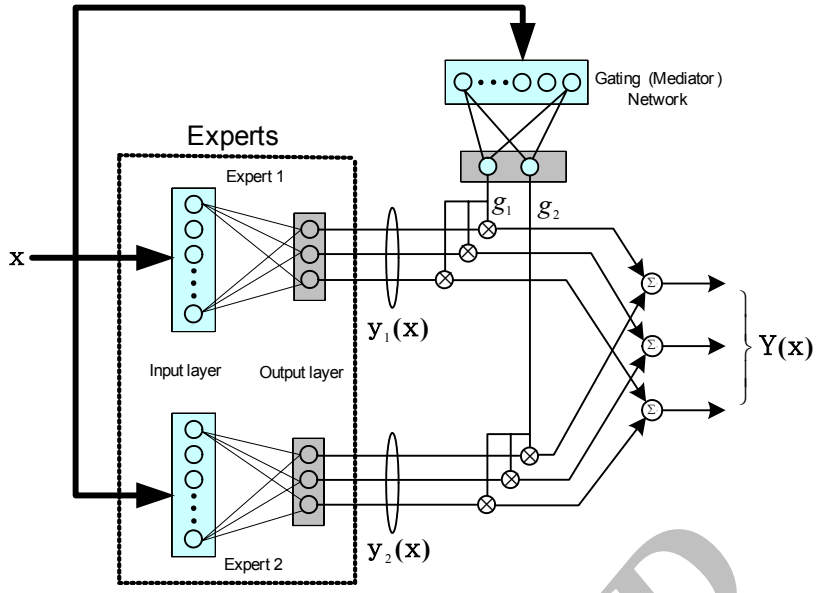
برای ترکیب نتایج طبقه‌بندها نیز قواعد مختلفی پیشنهاد شده است [۲۲] که ترکیب خطی نتایج یکی از متداول‌ترین آنها است. در این حالت، خروجی سیستم طبقه‌بندی مرکب یک ترکیب خطی از نظرات تمام طبقه‌بندهای پایه است. این روش با عنوان میانگین وزن‌دار نیز مطرح

این مقاله در تاریخ ۷ تیر ماه ۱۳۸۵ دریافت و در تاریخ ۳ تیر ماه ۱۳۸۶ بازنگری شد.
این پژوهش با پشتیبانی مرکز تحقیقات مخابرات ایران بر اساس قرارداد شماره ۵۰۰/۳۶۷۹/ت انجام شده است.

سید حسن نبوی کریزی، بخش مهندسی برق، دانشگاه تربیت مدرس، تهران، ایران، (email: nabavikarizi@yahoo.com).
احسان‌اله کبیر، بخش مهندسی برق، دانشگاه تربیت مدرس، تهران، ایران، (email: kabir@modares.ac.ir).

1. Ensemble (Committee) Learning
2. Base Classifiers
3. Diverse
4. Implicit
5. Explicit
6. Bagging Method
7. Boosting Methods
8. Penalty Methods

Archive of SID
www.SID.ir



شکل ۱: روش اختلاط خبره‌ها.

است. مشکل اصلی در این روش، تعیین وزن مربوط به نظر هر طبقه‌بند است. در ساده‌ترین حالت، که وزن تخصیص داده شده به نظر همه طبقه‌بندها یکسان باشد، روش را قاعده میانگین ساده می‌گویند. برای پیداکردن وزن‌های بهینه ترکیب خطی در مسائل رگرسیون، روش‌های مختلفی ارائه شده است اما بیشتر آنها برای مسائل طبقه‌بندی مناسب نیستند [۲۳]. استفاده از مدل رفتاری طبقه‌بندها روی نمونه‌های یادگیری، مثلاً استفاده از ماتریس سردرگمی طبقه‌بند، یک روش متداول برای پیداکردن وزن‌های ترکیب خطی در مسائل طبقه‌بندی است [۲۴].

ما در این تحقیق می‌خواهیم یک روش دومرحله‌ای برای ترکیب نتایج خروجی طبقه‌بندها ارائه کنیم که در مرحله اول آن از روش اختلاط خبره‌ها برای ایجاد گوناگونی بین طبقه‌بندهای پایه و در مرحله دوم آن از روش بهینه‌سازی‌گروه ذرات[1] (PSO) برای پیداکردن وزن‌های ترکیب خطی نظرات طبقه‌بندها استفاده می‌شود. با انجام آزمایش‌هایی بر روی چند مجموعه داده، کارآیی این روش دومرحله‌ای را نسبت به روش‌های یادگیری مستقل و روش اختلاط خبره‌ها نشان می‌دهیم.

ساختار بقیه مقاله به صورت زیر است. در بخش دوم روش اختلاط خبره‌ها تشریح می‌شود. در بخش سوم، روش پیشنهادی برای ترکیب نتایج طبقه‌بندها، به صورت یک روش دومرحله‌ای، معرفی می‌شود. بخش چهارم نتایج تجربی تحقیق را در بر دارد و بخش پنجم به نتیجه‌گیری و پیشنهاد برای ادامه کار می‌پردازد.

۲- اختلاط خبره‌ها

۲-۱ ساختار

ایده اختلاط خبره‌ها[2] (MOE)، بر مبنای استراتژی تقسیم‌کن و پیروزشو[3]، در سال ۱۹۹۱ توسط ژاکوب ارائه شد [۲۵]. این ساختار شامل تعدادی طبقه‌بند خطی (به عنوان خبره[4]) و یک طبقه‌بند خطی به عنوان میانجی[5] است. خبره‌ها در این ساختار همان طبقه‌بندهای پایه هستند. در این تحقیق از شبکه‌های عصبی پرسپترون یک‌لایه با تابع آستانه خطی به عنوان طبقه‌بندهای خبره و شبکه میانجی استفاده شده است. در این ساختار، یادگیری طبقه‌بندها از نوع یادگیری رقابتی است یعنی در هنگام یادگیری، طبقه‌بندی که خطای کمتری بر روی نمونه یادگیری دارد تشویق می‌شود. این ساختار برای حالتی که اختلاط خبره‌ها شامل دو خبره باشد در شکل ۱ نشان داده شده است.

شبکه میانجی دو وظیفه دارد. اول اینکه فضای ورودی را به صورت هوشمندانه بین خبره‌ها تقسیم کند و دوم اینکه با توجه به میزان هوشمندی هر خبره یک وزن به نظر آن خبره تخصیص دهد. هم‌زمان با یادگیری خبره‌ها، شبکه میانجی یاد می‌گیرد که چگونه وزن مربوط به نظر هر خبره را به صورت تابعی از الگوی ورودی و وزن‌های یادگیری‌اش محاسبه کند. در این ساختار، شبکه میانجی، با توجه به الگوی ورودی و وزن‌های یادگیری‌اش فقط یک وزن به نظر هر خبره تخصیص می‌دهد. ما در این تحقیق می‌خواهیم یک الگوریتم دومرحله‌ای بر مبنای این روش ارائه کنیم که در آن به نظر هر طبقه‌بند به تعداد کلاس‌های الگو وزن تخصیص داده شود. در ادامه پس از بررسی روش یادگیری اختلاط خبره‌ها به بیان روش پیشنهادی می‌پردازیم.

۲-۲ روش یادگیری اختلاط خبره‌ها

گفتیم یادگیری در این ساختار، از نوع یادگیری رقابتی است یعنی هر طبقه‌بندی که هنگام یادگیری خطای کمتری بر روی نمونه یادگیری داشته باشد بیشتر تشویق می‌شود [۲۵]. برای بررسی دقیق‌تر این موضوع فرض کنید $\Omega = \{\omega_1, \omega_2, \ldots, \omega_K\}$ مجموعه برچسب‌های کلاسی، $E = \{E_1, E_2, \ldots, E_L\}$ مجموعه خبره‌ها و $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ بردار الگوی ورودی باشند. اگر خروجی خبره iام برای بردار ورودی $\mathbf{x}$ به صورت $\mathbf{y}_i(\mathbf{x})$ و خروجی سیستم مرکب به صورت $\mathbf{y}(\mathbf{x})$ نشان داده شوند داریم

$$\mathbf{y}_i(\mathbf{x}) = \mathbf{w}_i^e . \mathbf{x} \qquad (۱)$$

$$\mathbf{Y}(\mathbf{x}) = \sum_{i=1}^{L} g_i \mathbf{y}_i(\mathbf{x}) \qquad (۲)$$

که در آن $\mathbf{w}_i^e$ بردار وزن خبره iام، عملگر $.$ نماد ضرب داخلی دو بردار $\mathbf{w}_i^e$ و $\mathbf{x}$، L تعداد خبره‌ها و g_i وزن تخصیص داده شده به نظر خبره

1. Particle Swarm Optimization
2. Mixture of Experts
3. Divide and Conquer
4. Expert
5. Mediator (Gating) Network

Archive of SID

www.SID.ir

iام (توسط شبکه میانجی) است. شبکه میانجی با توجه به احتمال تولید خروجی مطلوب توسط خبره iام، وزن g_i را به آن تخصیص می‌دهد. بنابراین می‌توان g_i را تخمینی از احتمال $p(E_i|\mathbf{x},\mathbf{w}_i^e)$ دانست. در این ساختار، تعداد نرون‌های لایه خروجی شبکه میانجی، برابر با تعداد خبره‌ها است. خروجی نرون iام شبکه میانجی را می‌توان به صورت (۳) بیان کرد

$$\mathbf{h_i}(\mathbf{x}) = \sum_{t=1}^{n} x_t \mathbf{w_{it}^g} \qquad i = 1,2,\ldots,L \ , \ t = 1,2,\ldots,n \qquad (۳)$$

که در آن $\mathbf{w_{it}^g}$ وزن اتصال بین نرون tام در لایه ورودی به نرون iام در لایه خروجی (برای شبکه میانجی) است. با توجه به مقدار h_iها، مقادیر g_iها از (۴) محاسبه می‌شوند

$$g_i(\mathbf{x}) = \frac{\exp(h_i(\mathbf{x}))}{\sum_{j=1}^{L}\exp(h_j(\mathbf{x}))} \qquad (۴)$$

با تعریف فوق برای g_i، مجموع وزن‌های تخصیص داده شده به خبره‌ها برابر یک خواهد بود. بزرگتربودن مقدار g_i برای یک خبره به معنی توانمندتربودن آن خبره برای تولید خروجی مطلوب است. در حالت مرزی، مقدار یک برای g_i به معنی این است که فقط با استفاده از خروجی خبره iام می‌توان کلاس الگوی ورودی را به درستی تعیین کرد و مقدار صفر برای g_i به معنی این است که نظر خبره iام در مورد کلاس الگوی ورودی نادرست است. در بقیه حالت‌ها کلاس الگوی ورودی به صورت ترکیب خطی از نظرات کلیه خبره‌های موجود در سیستم و توسط (۲) پیدا می‌شود. چنانکه (۴) نشان می‌دهد g_iها تابعی از الگوی ورودی $\mathbf{x}$ و وزن‌های یادگیری $\mathbf{w}^g$ هستند.

اصلاح وزن شبکه‌های خبره و میانجی به صورت (۵) و (۶) است

$$\Delta\mathbf{w}_i^e = \eta_e k_i(\mathbf{d}(\mathbf{x}) - \mathbf{y}_i(\mathbf{x})).\mathbf{x}^T \qquad (۵)$$

$$\Delta w_i^g = \eta_g (k_i - g_i).\mathbf{x}^T \qquad (۶)$$

در روابط فوق، $\mathbf{d}(\mathbf{x})$ بردار خروجی مطلوب برای الگوی ورودی $\mathbf{x}$، η_e و η_g به ترتیب نرخ یادگیری برای خبره iام و شبکه داور و k_i دارای تعریف زیر است

$$k_i = \frac{g_i \exp(\frac{-1}{2}(\mathbf{d}-\mathbf{y}_i)^T(\mathbf{d}-\mathbf{y}_i))}{\sum_{j=1}^{L} g_j \exp(\frac{-1}{2}(\mathbf{d}-\mathbf{y}_j)^T(\mathbf{d}-\mathbf{y}_j))} \qquad (۷)$$

چنانکه (۷) نشان می‌دهد، k_i مرتبط با مجموع مربعات خطای خبره iام است یعنی هر قدر مربع خطای خبره کمتر باشد تشویق بیشتری برای اصلاح وزن‌های یادگیری‌اش صورت می‌گیرد.

۳- روش پیشنهادی

دیدیم در روش اختلاط خبره‌ها به نظر هر طبقه‌بند یک وزن تعلق می‌گیرد. اگر بتوان به نظر هر طبقه‌بند در مورد هر یک از کلاس‌های الگو یک وزن تخصیص داد، انتظار می‌رود که دقت طبقه‌بندی بیشتر شود. ما در این تحقیق به دنبال این موضوع هستیم. چون مرحله دوم روش پیشنهادی از روش بهینه‌سازی گروه ذرات استفاده می‌کند ابتدا به معرفی این روش می‌پردازیم.

۳-۱ روش بهینه‌سازی گروه ذرات

روش بهینه‌سازی گروه ذرات، PSO، یک تکنیک بهینه‌سازی است که از رفتار جمعی پرندگان و ماهی‌ها در پیداکردن غذا الهام گرفته شده است [۲۶]. در این روش هر نامزد برای جواب مسأله، یک پرنده در فضای جستجو است که ذره[1] نام دارد. هر ذره دارای یک مقدار شایستگی است که توسط تابع شایستگی مسأله به دست می‌آید. پرنده‌ای که به غذا نزدیک‌تر است، شایستگی بیشتری دارد. این روش مانند اکثر روش‌های جستجوی تکاملی، با یک گروه از جواب‌های تصادفی (ذره‌ها)، جستجو را به شکل موازی شروع می‌کند و سپس برای یافتن جواب بهینه در فضای مسأله، با به هنگام کردن مکان ذره‌ها، به جستجو ادامه می‌دهد.

مکان هر ذره در هر تکرار متأثر از دو عامل است. اولین عامل، بهترین موقعیتی است که این ذره تاکنون کسب کرده است و دومین عامل بهترین موقعیتی است که تا به حال در همسایگی این ذره کسب شده است. مکان هر ذره بر اساس این دو عامل به‌هنگام می‌شود. به عبارت دیگر اعضای جمعیت از طرفی موظفند موقعیت خود را با تبعیت از بهترین عضو جمعیت، در یک شعاع همسایگی، تغییر دهند و از طرف دیگر لازم است بهترین موقعیتی را که تاکنون شخصاً تجربه کرده‌اند در حافظه خود نگهداری کرده و تمایلی نیز به سمت آن داشته باشند [۲۷]. با توجه به شعاع همسایگی، روش بهینه‌سازی گروه ذرات دو گونه است. اگر شعاع همسایگی به اندازه کل اعضای جمعیت باشد، روش را بهینه‌سازی گروه ذرات "بهترین کلی[2]" می‌گویند. در این حالت بهترین موقعیت کسب‌شده در همسایگی یک ذره، بهترین موقعیت کسب‌شده در بین کل اعضای جمعیت است. اگر شعاع همسایگی کمتر از اندازه کل جمعیت باشد، روش را بهینه‌سازی گروه ذرات "بهترین محلی[3]" می‌گویند. ما در این تحقیق از روش بهینه‌سازی گروه ذرات "بهترین کلی" استفاده کرده‌ایم.

در روش بهینه‌سازی گروه ذرات برای هر ذره i، یک موقعیت y_i و یک سرعت v_i در نظر گرفته می‌شود. اگر بعد مسأله بهینه‌سازی n فرض شود، بردار موقعیت و بردار سرعت ذره i به صورت زیر است

$$\mathbf{y}_i = [y_{i1}, y_{i2}, \ldots, y_{in}]$$
$$\mathbf{v}_i = [v_{i1}, v_{i2}, \ldots, v_{in}]$$

که در آن y_{id} و v_{id} به ترتیب معرف موقعیت مکانی و سرعت بعد dام ذره i است. سرعت و موقعیت این بعد ذره در تکرار $(t+1)$ به صورت (۸) و (۹) است

$$v_{id}(t+1) = \omega.v_{id}(t) + c_1.rand(P_{id}(t) - y_{id}(t)) + c_2.rand(\hat{P}_d(t) - y_{id}(t)) \qquad (۸)$$

$$y_{id}(t+1) = y_{id}(t) + v_{id}(t+1) \qquad (۹)$$

در روابط فوق، w وزن اینرسی در بازه $[0,1]$، c_1 و c_2 ضرایب یادگیری یا شتاب در بازه $[1,2]$ (c_1 را پارامتر شناختی[4] و c_2 را پارامتر اجتماعی[5] می‌گویند و معمولاً $c_1 = c_2$ در نظر گرفته می‌شود)، $rand$ عددی تصادفی با توزیع یکنواخت در بازه $[0,1]$، P_{id} بهترین موقعیت ذره iام در بعد dام تا کنون و $\hat{P}_d$ بهترین موقعیت در بین کل ذره‌ها برای بعد dام تا به حال است. ω برای ایجاد توازن لازم بین یافتن پاسخ کلی و محلی در (۸) وارد شده است. ثابت شده است که شرط همگرایی الگوریتم آنست که (۱۰) برقرار باشد [۲۸]

1. Particle
2. Global Best (G-best) PSO
3. Local Best (L-best) PSO
4. Cognitive Parameter
5. Social Parameter

Archive of SID

www.SID.ir

جدول ۱: انواع آزمایش‌های انجام‌شده در این تحقیق.

نوع آزمایش	روش آزمایش
M۱	طبقه‌بندهای پایه به صورت مستقل از یکدیگر تعلیم دیده و ترکیب‌خطی نظرات آنها با وزن‌های یکسان، خروجی سیستم مرکب را تشکیل می‌دهد.
M۲	طبقه‌بندهای پایه با ایده اختلاط خبره‌ها تعلیم دیده و ترکیب‌خطی نتایج آنها با وزن‌های یکسان، خروجی سیستم مرکب را تشکیل می‌دهد.
M۳	طبقه‌بندهای پایه با ایده اختلاط خبره‌ها تعلیم دیده و ترکیب‌خطی خروجی آنها با وزن‌هایی که شبکه‌میانجی می‌دهد، خروجی سیستم مرکب را تشکیل می‌دهد (روش اختلاط خبره‌ها).
M۴	طبقه‌بندهای پایه با ایده اختلاط خبره‌ها تعلیم دیده و ترکیب‌خطی خروجی آنها با وزن‌هایی که روش بهینه‌سازی گروه ذرات می‌دهد، خروجی سیستم مرکب را تشکیل می‌دهد (روش پیشنهادی).

$$0 \le \frac{\phi_1+\phi_2}{2} - 1 < \omega < 1 \qquad (10)$$

که در آن $\phi_1 = c_1.rand$ و $\phi_2 = c_2.rand$.

برای جلوگیری از واگرایی الگوریتم، معمولاً مقدار نهایی سرعت هر ذره به یک مقدار بیشینه محدود می‌شود، یعنی $v_{id} \in [-v_{\max}, v_{\max}]$. شرط خاتمه الگوریتم، همگرایی آن یا توقف بعد از تعداد معینی تکرار است [۲۹].

شایستگی هر ذره با استفاده از یک تابع برازندگی، $f(.)$، سنجیده می‌شود. این تابع برازندگی مربوط به مسأله مورد نظر بوده و هدف کمینه‌کردن آن است. بهترین موقعیت ذره iام در هر تکرار طبق (۱۱) به‌هنگام می‌شود

$$P_i(t) = \begin{cases} P_i(t-1) & if \quad f(y_i(t)) \ge f(P_i(t-1)) \\ y_i(t) & if \quad f(y_i(t)) < f(P_i(t-1)) \end{cases} \qquad (11)$$

معنی (۱۱) این است که اگر مقدار فعلی تابع برازندگی ذره iام، به ازای موقعیت فعلی y_i، از برازنده‌ترین مقدار قبلی آن کمتر باشد، همین موقعیت به عنوان بهترین موقعیت این ذره ثبت می‌شود و در غیر این صورت، بهترین موقعیت ذره، بهترین موقعیت قبلی آن باقی خواهد ماند.

۳-۲ بیان روش

گفتیم که ترکیب نتایج خروجی چند طبقه‌بند زمانی مؤثر واقع می‌شود که طبقه‌بندهای پایه با یکدیگر گوناگون بوده و قاعده ترکیب مناسبی برای تلفیق خروجی آنها به کار گرفته شود. روش پیشنهادی ما مبتنی بر این تفکر بوده و یک روش دومرحله‌ای است. مرحله اول این روش، با استفاده از تقسیم فضای ورودی بین طبقه‌بندهای پایه، توسط شبکه میانجی، یک مجموعه طبقه‌بند متفاوت ایجاد می‌کند. در مرحله دوم، با استفاده از عملکرد این خبره‌ها روی نمونه‌های آموزشی، وزن‌های بهینه ترکیب خطی پیدا شده و نتایج خروجی طبقه‌بندها بر اساس این وزن‌ها با یکدیگر تلفیق می‌شوند. از روش بهینه‌سازی گروه ذرات برای پیداکردن وزن‌های بهینه استفاده می‌شود. این روش، به نظر هر طبقه‌بند در مورد هر یک از کلاس‌های الگو یک وزن تخصیص می‌دهد.

در روش پیشنهادی، نرخ خطای طبقه‌بندی به عنوان تابع معیار تعریف می‌شود یعنی

$$f(\mathbf{w}) = \frac{E}{C+E} \qquad (12)$$

که در آن C و E به ترتیب مجموع تعداد نمونه‌هایی هستند که، توسط سیستم مرکب، درست و نادرست طبقه‌بندی شده‌اند. روش بهینه‌سازی گروه ذرات با کمینه‌کردن این تابع، وزن‌های ترکیب خطی را پیدا می‌کند.

۴- نتایج تجربی

به منظور ارزیابی روش پیشنهادی دو دسته آزمایش انجام دادیم. این آزمایش‌ها برای حالتی که سیستم مرکب شامل تعداد متفاوتی شبکه خطی (به عنوان طبقه‌بندهای پایه) باشد صورت گرفته است. اما در اینجا فقط برخی از نتایج آن ذکر می‌شود.

در هر دسته از آزمایش‌ها، چهار نوع آزمایش انجام شده است. این آزمایش‌ها در جدول ۱ تعریف شده‌اند.

برای آنکه تأثیر تصادفی‌بودن وزن‌های اولیه شبکه‌ها ناچیز باشد، هر آزمایش را ۱۰ بار تکرار کرده‌ایم و نتایج ارائه‌شده در این مقاله با متوسط‌گیری از این ۱۰ تکرار حاصل شده است. لازم به ذکر است که در این ۱۰ تکرار، نمونه‌های آموزشی و آزمایشی یکسان هستند.

پارامترهای استفاده‌شده در روش بهینه‌سازی گروه ذرات به صورت $w = 0.7298$، $c_1 = c_2 = 1.49445$ و $V_{\max} = 4$ هستند. این مقادیر، مقادیر متداولی هستند که در بیشتر مقالات مربوطه از جمله مراجع [۲۶] و [۳۰] استفاده شده‌اند. با سعی و خطا، ۵۰ مقدار مناسبی برای تعداد ذره‌ها در نظر گرفته شده است.

۴-۱ دسته اول آزمایش‌ها

در آزمایش‌های دسته اول از روش پیشنهادی برای پیداکردن مرز تصمیم یک مسأله دوکلاسه استفاده شده است. نمونه‌های هر دو کلاس توزیع گوسی داشته و به مقدار زیادی با یکدیگر تداخل دارند. نمونه‌های کلاس یک دارای میانگین $\mu_1 = [0,0]^T$ و واریانس $\sigma_1^2 = 1$ و نمونه‌های کلاس دو دارای میانگین $\mu_2 = [2,0]^T$ و واریانس $\sigma_2^2 = 4$ هستند. مجموعه آموزش شامل ۱۰۰۰ نمونه از هر کلاس و مجموعه آزمایش شامل ۱۶۰۰۰ نمونه از هر کلاس است [۳۱]. برخی نتایج چهار نوع آزمایش انجام‌شده بر روی این مجموعه داده در جدول ۲ نشان داده شده است. این نتایج به ازای $\eta_e = 0.25$ برای آموزش شبکه‌های عصبی پایه و $\eta_g = 0.01$ برای آموزش شبکه میانجی و با میانگین‌گیری از ۱۰ بار تکرار هر آزمایش، به دست آمده‌اند. همان‌طور که در بخش قبل نیز تأکید شد، طبقه‌بندهای پایه، شبکه‌های عصبی خطی هستند.

سطر متناظر با روش M۱ در این جدول نشان می‌دهد که اگر شبکه‌ها به صورت مستقل آموزش ببینند، همگی مرز تصمیم یکسانی را برای این مسأله دوکلاسه پیدا می‌کنند و ترکیب نتایج آنها بهبودی در نرخ بازشناسی سیستم مرکب حاصل نمی‌کند.

اختلاف قابل توجه بین میانگین نرخ بازشناسی طبقه‌بندهای پایه و سیستم مرکب در سطرهای متناظر با روش‌های M۲، M۳ و M۴ توانمندی روش اختلاط خبره‌ها در ایجاد شبکه‌هایی گوناگون با یکدیگر را نشان می‌دهد. واضح است که در این روش‌ها تخصیص وزن یکسان به شبکه‌های پایه نتیجه خوبی را برای سیستم مرکب حاصل نمی‌کند (سطر متناظر با روش M۲ را ببینید). بلکه باید با توجه به میزان هوشمندی هر شبکه، یک وزن مناسب به آن شبکه تخصیص داده شود. در روش M۳، شبکه میانجی وزن تخصیص داده شده به هر شبکه را تعیین می‌کند و در روش M۴، این وزن با روش بهینه‌سازی گروه ذرات (بر اساس کمینه‌کردن نرخ خطای طبقه‌بندی) تعیین می‌شود.

برای درک بیشتر، یک نمونه متعلق به کلاس ω_2، را که با روش M۲

Archive of SID
www.SID.ir

جدول ۲: میانگین نرخ بازشناسی شبکه‌های پایه و سیستم مرکب برای آزمایش‌های دسته اول.

تعداد شبکه‌ها	میانگین نرخ بازشناسی شبکه‌های پایه (٪)					میانگین نرخ بازشناسی سیستم مرکب (٪)	روش یادگیری
۳	۷۵٫۴۴	۷۵٫۴۴	۷۵٫۴۴			۷۵٫۴۴	M۱
	۶۳٫۱۲	۵۴٫۷۳	۶۹٫۵۱			۵۵٫۳۲	M۲
	۶۳٫۱۲	۵۴٫۷۳	۶۹٫۵۱			۸۰٫۳۳	M۳
	۶۳٫۱۲	۵۴٫۷۳	۶۹٫۵۱			۸۱٫۳۵	M۴
۵	۷۱٫۵۱	۷۱٫۵۱	۷۱٫۵۱	۷۱٫۵۱	۷۱٫۵۱	۷۵٫۵۱	M۱
	۶۴	۶۱٫۳۲	۶۷٫۹۵	۵۵٫۴۹	۶۶٫۱۶	۵۶٫۸۱	M۲
	۶۴	۶۱٫۳۲	۶۷٫۹۵	۵۵٫۴۹	۶۶٫۱۶	۸۰٫۷۶	M۳
	۶۴	۶۱٫۳۲	۶۷٫۹۵	۵۵٫۴۹	۶۶٫۱۶	۸۱٫۳۶	M۴

نادرست طبقه‌بندی شده ولی روش‌های M۳ و M۴ آنرا درست طبقه‌بندی کرده‌اند، بررسی می‌کنیم. برای این نمونه، ماتریس پروفایل تصمیم [۲۲] به صورت زیر است

$$DP(\mathrm{x})=\begin{bmatrix} ۰٫۹۹۹ & ۰٫۰۰۱ \\ ۰٫۳۷۲ & ۰٫۶۲۸ \\ ۰٫۹۹۵ & ۰٫۰۰۵ \end{bmatrix}$$

سطرهای این ماتریس، نظر طبقه‌بندهای E_1 تا E_3 و ستون‌های آن متناظر با کلاس‌های الگو هستند. به عنوان مثال سطر دوم این ماتریس بیانگر آنست که طبقه‌بند E_2 تعلق الگوی x به کلاس‌های ω_1 و ω_2 را به ترتیب ۰٫۳۷۲ و ۰٫۶۲۸ گزارش کرده است. در نگاه اول، با توجه به اینکه طبقه‌بندهای E_1 و E_3، با درجه تعلق بالایی، کلاس ω_1 را برای الگوی x پیشنهاد کرده‌اند به نظر می‌رسد که کلاس الگو ω_1 باشد در حالی که کلاس واقعی آن ω_2 است. در مورد این الگو، عملکرد روش‌های M۲ تا M۴ را بررسی می‌کنیم.

روش M۲: در این روش وزن یکسان به نظر همه خبره‌ها داده می‌شود و داریم

$$\mu_1(x)=\frac{(۰٫۹۹۹+۰٫۳۷۲+۰٫۹۹۵)}{۳}=۰٫۷۸۹$$

$$\mu_2(x)=\frac{(۰٫۰۰۱+۰٫۶۲۸+۰٫۰۰۵)}{۳}=۰٫۲۱۱$$

یعنی تعلق الگو به کلاس‌های ω_1 و ω_2 به ترتیب برابر ۰٫۷۸۹ و ۰٫۲۱۱ است. در این حالت الگوی x به صورت نادرست در کلاس ω_1 قرار می‌گیرد.

روش M۳: در این حالت بردار وزن تخصیص داده شده به خبره‌های E_1 تا E_3 به صورت زیر است

$$\mathrm{G}=[g_1,g_2,g_3]=[۰٫۰۰۰۲,۰٫۹۹۹,۰٫۰۰۰۸]$$

مفهوم این بردار وزن آنست که در مورد این الگو نظر طبقه‌بندهای E_1 و E_3 مهم نیست. وزن‌های بسیار ناچیز ۰٫۰۰۰۲ و ۰٫۰۰۰۸ که به این طبقه‌بندها تخصیص داده شده بیانگر این ادعاست. با این بردار وزن، مقدار عضویت الگو به کلاس‌های ω_1 و ω_2 به صورت زیر است

$$\mu_1(x)=۰٫۹۹۹\times۰٫۰۰۰۲+۰٫۳۷۲\times۰٫۹۹۹+۰٫۹۹۵\times۰٫۰۰۰۸=۰٫۳۷۳$$

$$\mu_2(x)=۰٫۰۰۱\times۰٫۰۰۰۲+۰٫۶۲۸\times۰٫۹۹۹+۰٫۰۰۵\times۰٫۰۰۰۸=۰٫۶۲۷$$

که در این صورت الگوی x به درستی در کلاس ω_2 طبقه‌بندی می‌شود.

روش M۴: در این روش ماتریس وزن تخصیص داده شده به نظر طبقه‌بندهای پایه با روش بهینه‌سازی گروه ذرات با تعداد تکرار ۸۰ به صورت زیر است

$$w_{PSO}=\begin{bmatrix} ۰٫۲۹۷ & ۰٫۰۸۷ \\ ۰٫۰۰۴ & ۰٫۹۹۸ \\ ۰٫۰۰۳ & ۰٫۹۸۶ \end{bmatrix}$$

که در آن به عنوان مثال وزن تخصیص داده شده به خبره دوم برای کلاس‌های ω_1 و ω_2 به ترتیب ۰٫۰۰۴ و ۰٫۹۹۸ است. با این ماتریس وزن داریم

$$\mu_1(x)=۰٫۹۹۹\times۰٫۲۹۷+۰٫۳۷۲\times۰٫۰۰۴+۰٫۹۹۵\times۰٫۰۰۳=۰٫۳۰۱$$

$$\mu_2(x)=۰٫۰۰۱\times۰٫۰۸۷+۰٫۶۲۸\times۰٫۹۹۸+۰٫۰۰۵\times۰٫۹۸۶=۰٫۶۳۲$$

چنانکه ملاحظه می‌شود در این روش نیز الگوی x به درستی در کلاس ω_2 طبقه‌بندی می‌شود. فاصله بین مقادیر تعلق به دو کلاس در این روش نسبت به روش M۳ بیشتر است.

این مثال، لزوم تخصیص وزن با توجه به ناحیه‌ای که طبقه‌بند پایه در آن خبره شده است (روش M۳) یا با توجه به برآورد هوشمندی آن با نمونه‌های آموزشی (روش M۴) را نشان می‌دهد.

برای آنکه تصور بهتری از میزان گوناگونی در طبقه‌بندهای پایه و اهمیت وزن‌دهی به نظر آنها داشته باشیم، در یکی از اجراهای برنامه تعداد خطاهای هر روش را شمارش کردیم. روی ۳۰۰۰۰ نمونه آزمایش، روش‌های M۱ تا M۴ به ترتیب ۷۳۶۹، ۱۳۳۵۷، ۵۹۴۰ و ۵۶۰۲ نمونه خطا داشتند. تعداد ۳۳۸۱ نمونه بودند که هر چهار روش در شناسایی آنها اشتباه کرده بودند. از ۱۳۳۵۷ نمونه خطای روش M۲، تعداد ۸۷۳۴ خطا با وزن‌دهی به روش اختلاط خبره‌ها (روش M۳) و ۹۱۱۹ خطا با وزن‌دهی به روش بهینه‌سازی گروه ذرات (روش M۴) رفع شدند. این بررسی ساده نشان می‌دهد که اولاً روش اختلاط خبره‌ها قادر به ایجاد طبقه‌بندهایی گوناگون است. ثانیاً وزن‌دهی به نظر طبقه‌بندهای پایه (به خاطر گوناگون‌بودن آنها) یک ضرورت است.

چنانکه نتایج مندرج در جدول نشان می‌دهند، کارآیی روش پیشنهادی از سایر روش‌ها بیشتر است.

در آزمایش‌های دسته اول تعداد نرون‌های لایه ورودی شامل ۳ نرون (که یکی مربوط به نرون بایاس) است و تعداد نرون‌های لایه خروجی به تعداد کلاس‌های الگو (یعنی دو تا) است. وزن‌های اولیه شبکه به صورت تصادفی انتخاب شده و نرخ یادگیری شبکه طبق رابطه $\eta=\eta_0(1-t/T)$ بود که در آن $\eta_0=۰٫۸$ نرخ اولیه یادگیری، t شماره تکرار جاری و $T=۱۰۰$ تعداد کل تکرارها است.

۴-۲ دسته دوم آزمایش‌ها

آزمایش‌های دسته دوم برای مسأله طبقه‌بندی چندکلاسه، بر روی مجموعه داده‌های تصاویر ماهواره‌ای، شیشه و یونسفر صورت گرفته است.

Archive of SID

www.SID.ir

جدول ۳: مشخصات مجموعه‌های داده.

مجموعه داده	تعداد نمونه‌ها	تعداد ویژگی‌ها	تعداد کلاس‌ها
تصاویر ماهواره‌ای	۶۴۳۵	۳۶	۶
شیشه	۲۱۴	۹	۶
یونسفر	۳۴۰	۳۳	۲

جدول ۴: میانگین نرخ بازشناسی طبقه‌بندهای پایه و سیستم مرکب بر روی مجموعه شش‌کلاسه تصاویر ماهواره‌ای.

تعداد خبره‌ها	میانگین نرخ بازشناسی طبقه‌بندهای پایه (٪)					میانگین نرخ بازشناسی سیستم مرکب (٪)	روش یادگیری
۳	۸۲٫۷۳	۸۲٫۸۱	۸۲٫۷۸			۸۲٫۷۹	M۱
	۵۸٫۵۴	۶۳٫۶۳	۴۵٫۸۸			۷۹٫۹۴	M۲
	۵۸٫۵۴	۶۳٫۶۳	۴۵٫۸۸			۸۲٫۹۶	M۳
	۵۸٫۵۴	۶۳٫۶۳	۴۵٫۸۸			۸۴٫۳۶	M۴
۵	۸۲٫۷۳	۸۲٫۷۵	۸۲٫۷۷	۸۲٫۷۷	۸۲٫۷۳	۸۲٫۷۷	M۱
	۲۳٫۸۸	۲۲٫۲۴	۴۵٫۸۸	۴۷٫۲۷	۵۲٫۵۳	۸۰٫۷۲	M۲
	۲۳٫۸۸	۲۲٫۲۴	۴۵٫۸۸	۴۷٫۲۷	۵۲٫۵۳	۸۲٫۹۰	M۳
	۲۳٫۸۸	۲۲٫۲۴	۴۵٫۸۸	۴۷٫۲۷	۵۲٫۵۳	۸۴٫۳۲	M۴

جدول ۵: میانگین نرخ بازشناسی طبقه‌بندهای پایه و سیستم مرکب بر روی مجموعه شش‌کلاسه شیشه.

تعداد شبکه‌ها	میانگین نرخ بازشناسی شبکه‌های پایه (٪)					میانگین نرخ بازشناسی سیستم مرکب (٪)	روش یادگیری
۳	۴۹٫۵۲	۴۹٫۵۲	۴۹٫۵۲			۴۹٫۵۲	M۱
	۴۹	۴۷٫۶۲	۵۰٫۴۸			۵۰٫۹۱	M۲
	۴۹	۴۷٫۶۲	۵۰٫۴۸			۵۴٫۳۸	M۳
	۴۹	۴۷٫۶۲	۵۰٫۴۸			۶۳٫۳۳	M۴
۵	۵۲٫۳۸	۵۱٫۹	۵۲٫۳۸	۵۱٫۹	۵۱٫۹	۵۱٫۹	M۱
	۵۳٫۳۳	۴۶٫۶۱	۴۸	۵۰	۴۹	۵۱٫۹	M۲
	۵۳٫۳۳	۴۶٫۶۱	۴۸	۵۰	۴۹	۵۵٫۲۳	M۳
	۵۳٫۳۳	۴۶٫۶۱	۴۸	۵۰	۴۹	۶۷٫۱۴	M۴

این مجموعه‌ها، از مجموعه‌های متداول برای استفاده در ترکیب طبقه‌بندها هستند [۳۲]. جدول ۳ مشخصات عمده این مجموعه‌ها را نشان می‌دهد. تعداد تکرار PSO در آزمایش‌های دسته دوم برابر ۱۰۰ است.

در این دسته از آزمایش‌ها تعداد نرون‌های لایه ورودی شامل $(K+1)$ نرون بود (که K بعد بردار ویژگی و یک مربوط به نرون بایاس است) و تعداد نرون‌های لایه خروجی به تعداد کلاس‌های الگو بود. وزن‌های اولیه شبکه به صورت تصادفی انتخاب شده و نرخ یادگیری شبکه طبق رابطه $\eta = \eta_0(1-t/T)$ بود که در آن $\eta_0 = ۰٫۸$ نرخ اولیه یادگیری، t شماره تکرار جاری و $T = ۱۰۰$ تعداد کل تکرارها است.

در مورد مجموعه داده تصاویر ماهواره‌ای چون تعداد نمونه‌ها زیاد است، ۶۶٪ نمونه‌ها به صورت تصادفی به عنوان نمونه‌های یادگیری (به منظور آموزش طبقه‌بندهای پایه و پیداکردن وزن‌های ترکیب خطی در روش‌های M۳ و M۴) و مابقی به عنوان نمونه‌های آزمایش در نظر گرفته شده‌اند. در مجموعه داده شیشه، چون تعداد نمونه‌ها کم است، از روش ارزیابی متقابل kبخشی[1] استفاده شده است. با انتخاب $k = ۱۰$، داده‌ها به ده بخش مساوی تقسیم شده و در هر آزمایش ۹ بخش برای آموزش شبکه‌ها و بخش باقیمانده برای آزمایش آنها استفاده شده است.

جدول ۴ میانگین نرخ بازشناسی طبقه‌بندهای پایه و سیستم مرکب بر روی مجموعه تصاویر ماهواره‌ای را نشان می‌دهد. برای این مجموعه داده شش کلاسه نیز همانند جدول ۲، کارآیی روش پیشنهادی نسبت به سه روش دیگر بیشتر است. اختلاف قابل توجه بین میانگین نرخ بازشناسی طبقه‌بندهای پایه و سیستم مرکب، نشان‌دهنده قابلیت روش پیشنهادی در ایجاد گوناگونی بین طبقه‌بندهای پایه است. دو نکته در مورد این مجموعه داده قابل توجه است. اول اینکه چون بعد بردار ویژگی نسبتاً زیاد است، نرخ بازشناسی طبقه‌بندهای پایه که به صورت مستقل از یکدیگر آموزش دیده‌اند اندکی با یکدیگر متفاوت است ولی باز هم نرخ بازشناسی سیستم مرکب از نرخ بهترین طبقه‌بند پایه بیشتر نیست. دوم آنکه در اینجا اختلاف بین نرخ بازشناسی روش پیشنهادی با روش M۳ نسبت به مجموعه داده قبلی بیشتر است. علت این امر بیشتربودن تعداد کلاس‌های الگو است. در اینجا شش کلاس الگو داریم در حالی که در مجموعه داده قبلی دو کلاس الگو داشتیم.

با افزایش تعداد کلاس‌های الگو، برتری روش پیشنهادی بر روش M۳ (روش استاندارد اختلاط خبره‌ها) بیشتر روشن می‌شود. نتایج مندرج در جدول ۵ برای مجموعه شش کلاسه شیشه نیز این موضوع را خاطرنشان می‌کنند.

در جدول ۵ میانگین نرخ بازشناسی طبقه‌بندهای پایه و سیستم مرکب برای مجموعه شش کلاسه شیشه نشان داده شده است. این جدول نیز همانند دو جدول قبل نشان می‌دهد که ترکیب چند شبکه‌خطی که به صورت مستقل از یکدیگر آموزش دیده‌اند، بهبودی در نرخ بازشناسی

1. K-Fold Cross Validation

Archive of SID

www.SID.ir

سیستم مرکب حاصل نمی‌کند و ترکیب نتایج آنها بهتر از بهترین شبکه پایه نیست. علت این امر اشتراک زیاد بین خطاهای شبکه‌های پایه با یکدیگر است. در حالیکه در سه روش دیگر همواره نرخ بازشناسی سیستم مرکب، بیشتر از نرخ بهترین شبکه پایه است.

بر روی مجموعه داده دو کلاسه یونسفر نیز آزمایش‌های فوق را انجام دادیم. برای این مجموعه داده میانگین نرخ بازشناسی سیستم مرکب در حالتی که شامل سه شبکه پایه بود برای روش‌های M۱ تا M۴ به ترتیب برابر ۹۱/۴۷، ۹۱/۱۸، ۹۸/۲۳ و ۹۷/۸۲ بود. در حالتی که سیستم مرکب شامل پنج شبکه پایه بود این نرخ‌ها به ترتیب ۸۷/۳۵، ۹۱/۴۷، ۹۸/۵۱ و ۹۷/۸۹ بود. این ارقام نشان می‌دهند که در مورد این مجموعه داده، کارآیی روش M۳ اندکی بیشتر از کارآیی روش پیشنهادی است. برای بررسی دقیق‌تر این موضوع به سراغ آزمون فرض t[1] رفتیم. آزمون فرض t برای سطح اطمینان $\alpha = ۰/۰۵$ تفاوت مهمی را بین روش پیشنهادی و روش M۳ نشان نداد. این به معنی آنست که دقت روش پیشنهادی روی این مجموعه داده نیز در حد دقت روش M۳ است.

نتایج آزمایش‌های گسترده روی مجموعه داده‌های مختلف برای ارزیابی بیشتر روش پیشنهادی نشان داد که:

– روش پیشنهادی برای زمانی که تعداد کلاس‌های الگو زیاد باشد کارآیی خوبی نسبت به روش استاندارد اختلاط خبره‌ها دارد.

– چون روش پیشنهادی، از مدل رفتاری طبقه‌بندها روی نمونه‌های آموزشی استفاده می‌کند، با افزایش تعداد نمونه‌های آموزشی کارآیی روش پیشنهادی بیشتر می‌شود.

– وقتی تعداد نمونه‌ها و تعداد کلاس‌های الگو کم باشند، کارآیی روش پیشنهادی در حد روش استاندارد اختلاط خبره‌ها است.

– زمان اجرای روش M۱ از سه روش دیگر کمتر است. روش‌های M۲ و M۳ تقریباً زمان اجرای یکسانی دارند و زمان اجرای روش پیشنهادی (روش M۴) در بدترین حالت ۱/۳۱ برابر روش M۱ است (بدترین حالت مربوط به مجموعه داده تصاویر ماهواره‌ای است). در بدترین حالت، با میانگین‌گیری از ۱۰ بار زمان اجرای روش‌ها بر روی مجموعه داده تصاویر ماهواره‌ای، زمان اجرای روش‌های M۲، M۳ و M۴ به ترتیب ۱/۲۲، ۱/۲۲ و ۱/۳۱ برابر روش M۱ به دست آمد.

– هر کدام از روش‌های M۱ تا M۴ شامل دو مرحله آموزش و آزمایش هستند. مقدار حافظه مصرفی برای مرحله آزمایش همه روش‌ها تقریباً یکسان است. در مرحله آموزش، مقدار حافظه مصرفی به تعداد طبقه‌بندهای پایه بستگی دارد. به عنوان مثال برای حالتی که سیستم مرکب شامل ۵ طبقه‌بند پایه است، مقدار حافظه مصرفی روش‌های M۲، M۳ به اندازه ۵/۴ برابر روش M۱ و مقدار حافظه مصرفی روش M۴ (با تعداد تکرار ۸۰ برای الگوریتم PSO)، حدود ۱/۴۵برابر روش M۱ است. توجه دارید که اگرچه برای اجرای الگوریتم PSO به حافظه نسبتاً زیادی نیاز است ولی اجرای این الگوریتم پس از آموزش طبقه‌بندهای پایه است و بنابراین می‌توان از بخشی از حافظه آزادشده در مرحله آموزش شبکه‌های پایه استفاده کرد.

۵- نتیجه‌گیری

در این مقاله یک روش دومرحله‌ای برای ترکیب نتایج طبقه‌بندها پیشنهاد شد که در مرحله اول آن، به روش اختلاط خبره‌ها یک مجموعه طبقه‌بند پایه با خطاهای متفاوت ایجاد می‌شود. در مرحله دوم، با استفاده از الگوریتم بهینه‌سازی گروه ذرات، وزن‌های بهینه برای ترکیب خطی نتایج طبقه‌بندهای پایه به دست آمده و بر اساس آن خروجی سیستم طبقه‌بندی مرکب شکل می‌گیرد. نتایج آزمایش‌های ما بر روی چهار مجموعه داده متداول، برای حالتی که طبقه‌بندهای پایه از نوع خطی هستند، نشان داد که روش پیشنهادی ما باعث افزایش نرخ بازشناسی سیستم طبقه‌بندی مرکب، نسبت به روش‌های یادگیری مستقل و اختلاط خبره‌ها می‌شود.

برخی نتایج آزمایش‌های ما نشان می‌دهند که کاهش تدریجی وزن اینرسی ω در روش بهینه‌سازی گروه ذرات (به صورت خطی) از مقدار اولیه ۰/۷۲۹۸ به مقدار نهایی ۰/۴ هم باعث افزایش سرعت همگرایی و هم باعث افزایش اندکی در نرخ بازشناسی صحیح سیستم می‌شود. به همین دلیل استفاده از مدل‌های دیگر PSO، نظیر Multi-phase PSO، Quantum PSO، PSO with Gaussian Mutation به عنوان یکی از موارد جالب برای ادامه این تحقیق خواهد بود.

همچنین نتایج آزمایش‌ها نشان داد که اضافه‌کردن یک ترم مومنتم به رابطه تصحیح وزن‌ها باعث افزایش سرعت همگرایی می‌شود. مطالعه و تحقیق بیشتر در این موضوع نیز برای ادامه کار پیشنهاد می‌شود.

مراجع

[1] F. Alimoglu and E. Alpaydin, "Combining multiple representations for pen-based handwritten digit recognition," *ELEKTRIC: Turkish J. of Electrical Engineering and Computer Sciences*, vol. 9, no. 1, pp. 1-12, 2001.

[2] V. Gunes and M. Menard, "Combination, cooperation and selection of classifiers: a state of the art," *IJPRAI*, vol. 17, no. 8, pp. 1303-1324, 2003.

[3] G. Giacinto, F. Roli, and L. Didaci, "Fusion of multiple classifier for intrusion detection in computer networks," *Pattern Recognition Letters*, vol. 24, no. 12, pp. 1795-1803, Aug. 2003.

[4] Y. Lu and C. L. Tan, "Combination of multiple classifiers using probabilistic dictionary and its application to postcode recognition," *Pattern Recognition*, vol. 35, no. 12, pp. 2823-2832, Dec. 2002.

[5] S. Gunter, Multiple Classifier Systems in Offline Cursive Handwriting Recognition, Ph.D. Tthesis, 2004.

[6] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270-2285, Dec. 2005.

[7] K. Chen and L. Wang, "Methods of combining multiple classifiers with different features and their applications to text-independent speaker identification," *IJPRAI*, vol. 11, no. 3, pp. 417-445, 1997.

[8] G. Rogova, "Combining the results of several neural network classifiers," *Neural Networks*, vol. 7, pp. 777-781, May 1994.

[9] L. Hansen and P. Salamon, "Neural network ensembles," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, no. 10, pp. 993-1001, Oct. 1990.

[10] A. Krogh and J. Vedelsby, "Neural network ensembles, cross validation, and active learning," in *Advances in Neural Information Processing Systems*, vol. 7, pp. 231-238, 1995.

[11] S. Hashem, B. Schmeiser, and Y. Yih, "Optimal linear combinations of neural networks: an overview," in *Proc. IEEE Int. Conf. on Neural Networks*, vol. 3, pp. 1507-1512, Orlando, Florida, 27 Jun.-2 Jul. 1994.

[12] R. Maclin and J. Shavlik, "Combining the predictions of multiple classifiers: using competitive learning to initialize neural networks," in *Proc. 14th Int. Joint Conf. on Artificial Intelligence*, pp. 524-530, Montreal, Canada, 1995.

[13] L. I. Kuncheva, M. Skurichina, and R. P. W. Duin, "An experimental study on diversity for bagging and boosting with linear classifiers," *Information Fusion*, vol. 3, no. 4, pp. 245-258, Dec. 2002.

[14] Y. Liu and X. Yao, "Ensemble learning via negative correlation," *Neural Networks*, vol. 12, no. 10, pp. 1399-1404, Dec. 1999.

[15] E. Bauer and R. Kohavi, "An empirical comparison of voting classification algorithms: bagging, boosting, and variants," *Machine Learning*, vol. 36, no. 1-2, pp. 105-142, Jul./Aug. 1999.

[16] W. Wang, P. Jones, and D. Partridge, "Diversity between neural networks and decision trees for building multiple classifier systems," *Lecture Notes in Computer Science*, vol. 1857, pp. 240-249, 2000.

1. t-Test

Archive of SID

www.SID.ir

[17] R. P. W. Duin and D. M. J. Tax, "Experiments with classifier combining rules," in *Proc. Int. Workshop on Multiple Classifier Systems, LNCS 1857*, vol. ???, pp. 16-29, ???. 2000.
[18] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123-140, Aug. 1996.
[19] Y. Raviv and N. Intrator, "Bootstrapping with noise: an effective regularization technique," *Connection Science*, vol. 8, pp. 355-372, 1996.
[20] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proc. 13th Int. Conf. on Machine Learning*, pp. 148-156, 1996.
[21] H. Drucker, C. Cortes, L. D. Jackel, Y. LeCun, and V. Vapnik, "Boosting and other ensemble methods," *Neural Computation*, vol. 6, pp. 1289-1301, Nov. 1994.
[۲۲] س. ح. نبوی کریزی و ا. کبیر، "ترکیب طبقه‌بندها: ایجاد گوناگونی و قواعد ترکیب،" *مجله علوم و مهندسی کامپیوتر، نشریه علمی پژوهشی انجمن کامپیوتر ایران*، جلد ۳، شماره ۳، صص. ۹۵-۱۰۷، پاییز ۱۳۸۴.
[23] N. Ueda, "Optimal linear combination of neural networks for improving classification performance," *IEEE Trans. on Pattern Anal. Mach. Intell.*, vol. 22, no. 2, pp. 207-215, Feb. 2000.
[24] C. L. Liu, "Classifier combination based on confidence transformation," *Pattern Recognition*, vol. 38, no. 1, pp. 11-28, Jan. 2005.
[25] M. Jordan and R. Jacobs, Modular and Hierarchical Learning Systems. The Handbook of Brain Theory and Neural Networks, MIT Press, Cambridge, MA, 1995.
[26] J. Kennedy and R. C. Eberhart, "Particle swarm optimization," in *Proc. of IEEE Int. Conf. on Neural Networks (ICNN)*, vol. 4, pp. 1942-1948, 1995.
[۲۷] س. ح. ظهیری، "بررسی تئوریک عملکرد طبقه‌بندی‌کننده‌های گروه ذرات،" *مجموعه مقالات یازدهمین کنفرانس بین‌المللی کامپیوتر انجمن کامپیوتر ایران*، جلد اول، صص. ۱۰-۳، بهمن ۱۳۸۴.
[28] I. C. Trelea, "The particle swarm optimization algorithm: convergence analysis and parameter selection," *Information Processing Letters*, vol. 85, no. 6, pp. 317-325, Mar. 2003.
[۲۹] م. رستمی شهر بابکی و ح. نظام آبادی‌پور، "انتخاب ویژگی در طبقه‌بندی معنایی تصاویر با استفاده از الگوریتم PSO،" *مجموعه مقالات یازدهمین کنفرانس بین‌المللی کامپیوتر انجمن کامپیوتر ایران*، جلد اول، صص. ۲۷۶-۲۶۹، بهمن ۱۳۸۴.
[30] J. Kennedy and R. C. Eberhart, *Swarm Intelligence*, Morgan Kaufmann Publishers, 2001.
[31] Y. Liu, X. Yao, Q. Zhao, and T. Higuchi, "An experimental comparison of ensemble learning methods on decision boundaries," in *Proc. IEEE International Joint Conf. on Neural Networks, IJCNN*, vol. 1, pp. 221-226, Honolulu, US, May 2002.
[32] http://www.ics.uci.edu/~mlearn/databases/ and www.dice.ucl.ac.be/neural-nets/research/projects/ELENA/database/.

سید حسن نبوی کریزی در مهر ماه ۱۳۶۸ پس از اتمام تحصیلات متوسطه وارد دانشکده مهندسی دانشگاه فردوسی مشهد شد و در آذر سال ۱۳۷۳ با اخذ مدرک کارشناسی در مهندسی الکترونیک از این دانشگاه فارغ‌التحصیل شد. وی در مهر ماه سال ۱۳۷۴ دوره کارشناسی ارشد خود را در دانشکده مهندسی دانشگاه تربیت مدرس آغاز کرد و در شهریور ۱۳۷۷ این دوره را به اتمام رسانید. او مدت ۳ سال به عنوان مدرس آموزشکده فنی مشهد مشغول به تدریس بود. وی از مهر ماه سال ۱۳۸۱ دوره دکتری مهندسی الکترونیک را در دانشگاه تربیت مدرس آغاز کرد و در شهریور ۱۳۸۵ از رساله خود دفاع کرد. او اکنون به عنوان مدرس آموزشکده فنی مشهد مشغول به تدریس است. زمینه‌های تحقیقاتی مورد علاقه وی عبارتند از: ترکیب طبقه‌بندها، بازشناسی الگو، پردازش تصویر.

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

Archive of SID

www.SID.ir