یک روش یادگیری جدید برای افزایش کارایی سیستم طبقه‌بندی مرکب

سیدحسن نبوی کریزی و احسان‌اله کبیر

چکیده: یادگیری دسته‌جمعی یک رویکرد مؤثر در یادگیری ماشینی است که به منظور بهبود کارایی سیستم بازشناسی الگو استفاده می‌شود. برای آنکه این نوع یادگیری مفید واقع شود باید خطاهای طبقه‌بندهای پایه با یکدیگر متفاوت باشند. راهکارهای ایجاد تفاوت در خطا، به دو گروه ضمنی و صریح تقسیم می‌شوند. در این تحقیق روش جدیدی از نوع صریح برای ایجاد گوناگونی در طبقه‌بندهای یک سیستم مرکب ارائه می‌شود. در این روش، معیار تازه‌ای از گوناگونی در فرایند یادگیری سیستم مرکب به کار گرفته می‌شود. در روش پیشنهادی، شباهت بین خطای هر طبقه‌بند با طبقه‌بندهای دیگر به صورت یک مؤلفه در تابع خطای آن طبقه‌بند منظور شده و در الگوریتم یادگیری آن ایفای نقش می‌کند. نتایج آزمایش‌های ما بر روی چند مجموعه داده متداول، برای حالتی که طبقه‌بندهای پایه از نوع شبکه‌های عصبی باشند، نشان می‌دهند که روش پیشنهادی ما باعث افزایش کارایی سیستم طبقه‌بندی مرکب نسبت به روش‌های مشابه آن می‌شود.

کلید واژه: **ترکیب طبقه‌بندها، یادگیری دسته‌جمعی، گوناگونی در خطا، همبستگی منفی، شبکه عصبی.**

۱- مقدمه

استفاده از یادگیری دسته‌جمعی[1]، یک رویکرد مؤثر در یادگیری ماشینی است که در سالهای اخیر محققین زیادی به آن پرداخته‌اند [۱]. در این نوع یادگیری، به منظور بهبود دقت یادگیری، نتایج چند طبقه‌بند با یکدیگر ترکیب می‌شوند. به این طبقه‌بندها، طبقه‌بندهای پایه[2] گفته می‌شود. شبکه‌های عصبی متداول‌ترین انتخاب برای طبقه‌بندهای پایه هستند [۲].

نتایج تئوری [۳] و [۴] و تجربی [۵] تا [۷] نشان می‌دهند که زمانی یادگیری دسته‌جمعی از یادگیری بهترین طبقه‌بند پایه بهتر است که طبقه‌بندهای پایه دارای کارایی قابل قبول بوده و گوناگون در خطا باشند. 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 و به همین علت گوناگونی در خطا[3]، از نکات اساسی در موفقیت یک سیستم طبقه‌بندی مرکب است [۸]. برای ایجاد گوناگونی در خطا، روشهای متعددی پیشنهاد شده است. این روشها به دو دسته کلی ضمنی[4] و صریح[5] تقسیم می‌شوند.

روشهای ضمنی، با تغییرات ضمنی در فرآیند یادگیری طبقه‌بندهای پایه، سعی در گوناگون کردن آنها دارند. در این روشها در طی یادگیری، هیچ معیاری از گوناگونی بررسی نمی‌شود و لذا هیچگونه تضمینی برای متفاوت‌شدن طبقه‌بندها وجود ندارد. استفاده از مجموعه‌های یادگیری متفاوت [۹]، ساختارهای مختلف [۱۰]، شرایط اولیه گوناگون، بازنمایی‌های متفاوت برای بیان الگو [۱۱]، تقسیم تصادفی فضای ویژگی بین طبقه‌بندهای پایه [۱۲]، انتخاب تصادفی نمونه‌های آموزشی هر طبقه‌بند [۱۳]، افزودن نویز به داده‌های ورودی [۱۴] و انجام تبدیلات غیر خطی بر روی داده‌های ورودی [۱۵] 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 متفاوت می‌سازند. این روشها در فرایند یادگیری طبقه‌بندها، معیاری از گوناگونی را اعمال کرده و بر اساس آن مسیر یادگیری طبقه‌بندها را در فضای یادگیری، تغییر می‌دهند تا طبقه‌بندهایی گوناگون در خطا ایجاد شوند.

روشهای تقویتی[7] [۱۶] و روشهای جریمه‌ای[8] [۱۷] 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 است. روش‌های جریمه‌ای، با بکارگیری یک مؤلفه جریمه در تابع خطای طبقه‌بندهای پایه، به ایجاد گوناگونی در آنها می‌پردازند. در این روشها، تعریف مؤلفه جریمه نقشی اساسی در ایجاد گوناگونی دارد.

ما در این تحقیق با استفاده از یک مؤلفه جریمه جدید، روشی برای ایجاد گوناگونی در طبقه‌بندهای پایه ارائه می‌کنیم و با انجام آزمایش‌هایی بر روی چند مجموعه داده، کاراتر بودن این روش را نسبت به روشهای یادگیری مستقل[9] و یادگیری همبستگی‌منفی[10] نشان می‌دهیم.

روشهای ضمنی عمدتاً با انجام عملیاتی بر روی نمونه‌های آموزشی سعی می‌کنند طبقه‌بندها را گوناگون کنند و در مرحله یادگیری معیاری از مقدار گوناگونی را در نظر نمی‌گیرند اما روشهای صریح به داده‌های آموزشی کاری ندارند بلکه مستقیماً در فرایند یادگیری، مقدار گوناگونی را

این مقاله در تاریخ ۳ مهر ماه ۱۳۸۴ دریافت و در تاریخ ۳۱ اردیبهشت ماه ۱۳۸۵ بازنگری شد. این تحقیق از پشتیبانی مالی مرکز تحقیقات مخابرات ایران براساس قرارداد شماره ۵۰۰/۳۶۷۹ت ۱۳۵۷ برخوردار بوده است.

سید حسن نبوی کریزی، بخش مهندسی برق، دانشگاه تربیت مدرس، تهران، کدپستی ۱۴۳-۱۴۱۱۵، ایران (email: hnabavi@modares.ac.ir).

احسان‌اله کبیر، بخش مهندسی برق، دانشگاه تربیت مدرس، تهران، کدپستی ۱۴۳-۱۴۱۱۵، ایران (email: kabir@modares.ac.ir).

1. Ensemble (Committee) Learning
2. Base Classifiers
3. Error Diversity
4. Implicit
5. Explicit
6. Bagging Method
7. Boosting Methods
8. Penalty Methods
9. Independent Learning
10. Negative Correlation Learning

Archive of SID

www.SID.ir

به صورت مؤلفه‌ای از تابع خطا در نظر گرفته و بر اساس آن مسیر یادگیری طبقه‌بندها را (در فضای یادگیری) متفاوت می‌سازند [۱۸].

در ادامه مقاله، در بخش دوم ضمن بیان فلسفه استفاده از مؤلفه جریمه، روش‌های جریمه متداول بیان می‌شوند. در بخش سوم، با معرفی یک مؤلفه جریمه‌ای جدید، روش پیشنهادی تشریح می‌شود. بخش چهارم نتایج تجربی تحقیق را در بر دارد و بخش پنجم به نتیجه‌گیری و پیشنهاد برای ادامه کار می‌پردازد.

۲- روشهای جریمه‌ای برای ایجاد گوناگونی

۲-۱ فلسفه استفاده از روشهای جریمه

اگر معیار خطا برای یک سیستم مرکب به صورت مجموع مربعات خطا در نظر گرفته شود، می‌توان خطای سیستم مرکب را به سه مؤلفه بایاس، واریانس و کواریانس تجزیه کرد [۱۹]. گرچه تئوری بایاس، واریانس و کواریانس ابتدا برای مسائل رگرسیون مطرح شد اما در مرجع [۲۰] و [۲۱] این تئوری برای مسائل طبقه‌بندی نیز تعمیم داده شده است. برای حالتی که سیستم طبقه‌بندی مرکب شامل L شبکه عصبی و خروجی آن میانگین خروجی‌های طبقه‌بندهای پایه فرض شود برای بردار ورودی x_p داریم

$$f_{ens}(x_p)=\frac{1}{L}\sum_{i=1}^{L}f_i(x_p) \qquad (1)$$

که در آن $f_{ens}(x_p)$ و $f_i(x_p)$ به ترتیب خروجی سیستم مرکب و خروجی شبکه iام برای بردار ورودی x_p هستند. اگر خطای سیستم مرکب را با E_{ens} نشان دهیم داریم

$$E_{ens}=E\{\frac{1}{L}\overline{\mathrm{Var}}+(1-\frac{1}{L})\overline{\mathrm{Cov}}+\overline{\mathrm{Bias}}^2\} \qquad (2)$$

که در آن

$$\overline{\mathrm{Bias}}=\frac{1}{L}\sum_{i=1}^{L}\mathrm{Bias}_i$$

$$\overline{\mathrm{Var}}=\frac{1}{L}\sum_{i=1}^{L}\mathrm{Var}_i$$

$$\overline{\mathrm{Cov}}=\frac{1}{L(L-1)}\sum_{i,j=1,i\neq j}^{L}\mathrm{Cov}_{i,j}$$

در (۲) عملگر $E\{\}$ امید ریاضی، Bias_i و Var_i به ترتیب بایاس و واریانس برای شبکه iام و $\mathrm{Cov}_{i,j}$ کواریانس بین خطای شبکه‌های i و j است که به صورت (۳) تا (۵) تعریف می‌شوند

$$\mathrm{Bias}_i=\frac{1}{N}\sum_{p=1}^{N}(E\{f_i(x_p)\}-d(x_p)) \qquad (3)$$

$$\mathrm{Var}_i=\frac{1}{N}\sum_{p=1}^{N}E\{(f_i(x_p)-E\{f_i(x_p)\})^2\} \qquad (4)$$

$$\mathrm{Cov}_{i,j}=\frac{1}{N.N}\sum_{P=1}^{N}E\{(f_i(x_p)-E\{f_i(x_p)\}).(f_j(x_p)-E\{f_j(x_p)\})\} \qquad (5)$$

که در آن $d(x_p)$ خروجی مطلوب برای بردار ورودی x_p و N تعداد کل نمونه‌های آموزشی است.

چنانکه روابط فوق نشان می‌دهند، بایاس معیاری از نزدیکی پاسخ شبکه به پاسخ واقعی و واریانس معیاری از پایداری پاسخ شبکه بر روی نمونه‌هایی است که اندکی با نمونه‌های آموزشی متفاوت‌اند. این دو مؤلفه معمولاً در جهت عکس یکدیگر رفتار می‌کنند، یعنی افزایش (کاهش) یکی باعث کاهش (افزایش) دیگری می‌شود. بنابراین نمی‌توان هر دوی آنها را به صورت همزمان کمینه کرد.

اگر همه شبکه‌ها دارای ساختار یکسان باشند، (۲) به صورت (۶) ساده خواهد شد

$$E_{ens}=E\{\frac{1}{L}\mathrm{Var}+(1-\frac{1}{L})\mathrm{Cov}+\mathrm{Bias}^2\} \qquad (6)$$

مقادیر بایاس و واریانس برای هر طبقه‌بند مربوط به خود آن طبقه‌بند است و به سایر طبقه‌بندهای سیستم مرکب بستگی ندارد، اما کواریانس مرتبط با رابطه بین خطای طبقه‌بندهای موجود در سیستم مرکب بوده و در حقیقت معیاری از همبستگی بین خطاهای طبقه‌بندهای پایه است. هر قدر خطاهای طبقه‌بندهای پایه بیشتر به هم شبیه باشند، مقدار مؤلفه کواریانس بیشتر خواهد بود. این مؤلفه بیانگر رابطه متقابل بین خطای طبقه‌بندهای پایه است. با توجه به اینکه رفتار دو مؤلفه بایاس و واریانس در جهت عکس یکدیگر است و کاهش یکی باعث افزایش دیگری می‌شود، یک روش مؤثر برای کمینه‌کردن خطای سیستم مرکب، کاهش مؤلفه کواریانس است، بدون آنکه مؤلفه‌های بایاس و واریانس افزایش پیدا کنند. رابطه (۶) نشان می‌دهد که با کاهش کواریانسِ خطای شبکه‌های پایه، می‌توان خطای سیستم مرکب را کاهش داد. کواریانس را می‌توان از طریق کاهش همبستگی در موارد خطای طبقه‌بندهای پایه، کاهش داد. یک روش مؤثر برای این کار استفاده از یک مؤلفه جریمه، R، در تابع خطای طبقه‌بندهای پایه است [۲۱]. شکل کلی تابع خطای طبقه‌بند iام در روشهای جریمه‌ای به صورت (۷) است

$$e_i(x_p)=\frac{1}{2}(f_i(x_p)-d(x_p))^2+\lambda R(x_p) \qquad (7)$$

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

۲-۲ روشهای جریمه‌ای متداول

چنانکه بیان شد روشهای جریمه‌ای، با اضافه کردن یک مؤلفه جریمه به تابع خطای طبقه‌بندهای پایه مسیر یادگیری آنها را در فضای یادگیری متفاوت می‌سازند. مؤلفه جریمه معمولاً یک مؤلفه مبتنی بر همبستگی[1] خطاهای طبقه‌بندهای پایه است.

روزن در [۲۲] برای گوناگون کردن شبکه‌ها یک مؤلفه جریمه به صورت زیر معرفی کرده است

$$R(x_p)=\sum_{j=1}^{i-1}c(j,i).(d(x_p)-f_i(x_p)).(d(x_p)-f_j(x_p)) \qquad (8)$$

در این روش، شبکه‌ها به صورت متوالی آموزش می‌بینند و هر شبکه هنگام آموزش نه تنها انحراف پاسخش از مقدار واقعی بلکه شباهت خطاهایش با شبکه‌های قبلش را نیز حداقل می‌کند. در این حالت حاصلضرب بایاس شبکه‌های i و j به عنوان مؤلفه جریمه در نظر گرفته می‌شود. ضریب c مشخص می‌کند که کدام شبکه‌ها بایستی در خطا غیر همبسته شوند. برای حالتی که هر شبکه بخواهد فقط با شبکه قبلش

1. Correlation

Archive of SID
www.SID.ir

جدول ۱: مشخصات مجموعه‌های داده

مجموعه داده	تعداد نمونه‌ها	تعداد ویژگی‌ها	تعداد کلاس‌ها
یونوسفر	۳۵۱	۳۴	۲
تصاویر ماهواره‌ای	۶۴۳۵	۳۶	۶
سونار	۲۰۸	۶۰	۲

غیر همبسته شود، ضریب c به صورت زیر خواهد بود

$$c(j,i)=\begin{cases}1 & if\ \ i=j-1\ \ and\ \ i\ \ is\ \ even\\ 0 & otherwise\end{cases} \quad (9)$$

اشکال این روش این است که چون شبکه‌ها به صورت متوالی آموزش می‌بینند، فرایند یادگیری سیستم مرکب وقت‌گیر است.

یانگ لیو در ادامه کارهای روزن، مؤلفه جریمه‌ای پیشنهاد کرد که با آن همه شبکه‌ها به صورت همزمان آموزش می‌بینند [۲۳]. این روش با عنوان یادگیری همبستگی منفی در کاربردهای مختلفی نظیر رگرسیون [۲۴]، تخمین سریهای زمانی [۲۳] و تقریب تابع چند متغیره [۲۵] و طبقه‌بندی [۸] به کار گرفته شده و در مقایسه با دیگر روشهای یادگیری دسته‌جمعی کارایی بهتری داشته است.

مؤلفه جریمه در این روش به صورت زیر است

$$R_i(x_p)=(f_i(x_p)-f_{ens}(x_p)).\sum_{j=1,j\neq i}^{L}(f_j(x_p)-f_{ens}(x_p)) \quad (10)$$

که در آن، $f_{ens}(x_p)$ خروجی سیستم مرکب برای ورودی x_p است. با این مؤلفه جریمه، هر شبکه در طی یادگیری نه تنها خطایش را نسبت به خروجی مطلوب کاهش می‌دهد، بلکه تنوعش با شبکه‌های دیگر را نیز افزایش می‌دهد و پارامترهای یادگیریش را بر اساس کمینه‌شدن خطا و بیشینه‌شدن گوناگونی تنظیم می‌کند.

اگر طبق (۱) خروجی سیستم مرکب، میانگین ساده خروجی شبکه‌های پایه در نظر گرفته شود، (۱۰) به صورت (۱۱) ساده خواهد شد

$$R_i(x_p)=-(f_i(x_p)-f_{ens}(x_p))^2 \quad (11)$$

بنابراین تابع خطای شبکه iام برای ورودی x_p به صورت (۱۲) خواهد بود

$$e_i(x_p)=\frac{1}{2}(f_i(x_p)-d(x_p))^2-\lambda(f_i(x_p)-f_{ens}(x_p))^2 \quad (12)$$

با توجه به رابطه فوق خطای شبکه iام شامل دو مؤلفه است. انحراف پاسخ شبکه از مقدار واقعی و انحراف آن از پاسخ سیستم مرکب. دور بودن پاسخ شبکه از میانگین پاسخ طبقه‌بندها باعث کاهش خطا و تأثیر آن در تصحیح وزنهای شبکه خواهد شد.

۳- روش پیشنهادی برای ایجاد گوناگونی

چنانکه بیان شد، در روش روزن شبکه‌ها به صورت متوالی آموزش می‌بینند و لذا فرایند یادگیری سیستم مرکب وقت‌گیر است. در روش یادگیری همبستگی منفی، گرچه همه شبکه‌ها به صورت همزمان آموزش می‌بینند، چنانچه همه آنها پاسخ یکسان ولی نادرست به الگوی ورودی x_p بدهند، اثر مؤلفه جریمه صفر خواهد شد. برای رفع این مشکل مؤلفه جریمه جدیدی به صورت زیر پیشنهاد می‌کنیم

$$R_i(x_p)=(f_i(x_p)-d(x_p)).\sum_{j=1,j\neq i}^{L}(f_j(x_p)-d(x_p)) \quad (13)$$

با این تعریف برای مؤلفه جریمه، شباهت بین خطای شبکه iام با خطاهای سایر شبکه‌ها یک مؤلفه در تابع خطای این شبکه به حساب می‌آید. در این حالت اگر همه طبقه‌بندها به الگوی ورودی x_p پاسخ یکسان ولی نادرست بدهند، مؤلفه جریمه غیر صفر بوده و باعث افزایش خطا می‌شود و لذا تصحیح وزن‌های شبکه با شدت بیشتری صورت خواهد گرفت.

با جایگذاری (۱۳) در (۷) داریم

$$e_i(x_p)=\frac{1}{2}(f_i(x_p)-d(x_p))^2+\lambda(f_i(x_p)-d(x_p)).\sum_{j=1,j\neq i}^{L}(f_j(x_p)-d(x_p)) \quad (14)$$

اگر w_{kl} وزن اتصال بین نرون lام در لایه مخفی و نرون kام در لایه خروجی باشد، مشتق مقدار خطای حاصل در نرون kام نسبت به وزن اتصال w_{kl} به صورت (۱۵) خواهد بود

$$\frac{\partial e_k(x_p)}{\partial w_{kl}}=[(1-2\lambda)f_k(x_p)+L\lambda\overline{f}(x_p)+(2\lambda-L\lambda-1)d(x_p)].\frac{\partial g(net_k(x_p))}{\partial(net_k(x_p))}.f_l^h(x_p) \quad (15)$$

که در آن، $net_k(x_p)$ و $f_l^h(x_p)$ به ترتیب خروجی نرون kام در لایه خروجی قبل از عبور از تابع آستانه و خروجی نرون lام در لایه مخفی پس از عبور از تابع آستانه، برای بردار ورودی x_p، هستند. $g(\cdot)$ یک تابع آستانه است. اگر تابع آستانه را به شکل سیگموئید در نظر بگیریم، داریم

$$\frac{\partial g(net_k(x_p))}{\partial(net_k(x_p))}=f_k(x_p).(1-f_k(x_p)) \quad (16)$$

بنابراین مقدار تغییر w_{kl} با (۱۷) بیان خواهد شد

$$\Delta w_{kl}=-\eta.\frac{\partial e_k(x_p)}{\partial w_{kl}} \quad (17)$$

که در آن η نرخ یادگیری شبکه است.

۴- یافته‌های تجربی

۴-۱ مجموعه داده‌ها

آزمایش‌های انجام‌شده بر روی مجموعه‌های داده یونوسفر[1]، تصاویر ماهواره‌ای[2] و سونار[3] صورت گرفته است. این مجموعه‌ها، از متداول‌ترین مجموعه‌های داده برای استفاده در ترکیب طبقه‌بندها هستند [۲۶]. در مجموعه داده تصاویر ماهواره‌ای چون تعداد نمونه‌ها زیاد است، ۷۵٪ نمونه‌ها به صورت تصادفی به عنوان نمونه‌های یادگیری و ۲۵٪ نمونه‌ها به عنوان نمونه‌های آزمایش در نظر گرفته شده‌اند. در دو مجموعه دیگر، چون تعداد نمونه‌ها کم است، از روش ارزیابی متقابل kبخشی[4] استفاده شده است. با انتخاب $k=6$ داده‌ها به شش بخش مساوی تقسیم شده و در هر آزمایش ۵ بخش برای آموزش شبکه‌ها و بخش باقیمانده برای آزمایش آنها استفاده شده است. جدول ۱ مشخصات عمده این مجموعه‌های داده را نشان می‌دهد.

1. Ionosphere
2. Satimage
3. Sonar
4. K-Fold Cross Validation

Archive of SID

www.SID.ir

جدول ۲: میانگین نرخ بازشناسی بر روی مجموعه داده یونوسفر: تعداد نرون‌های لایه مخفی دو عدد و تعداد دوره‌ها در هر آزمایش ۲۵۰ بوده است. به ازای هر مقدار λ، آزمایش ۱۵ بار تکرار شده و میانگین نتایج در جدول آمده است.

تعداد شبکه‌ها	λ^*	میانگین نرخ بازشناسی شبکه‌های پایه (٪)						میانگین نرخ بازشناسی سیستم مرکب (٪)	روش یادگیری
۶	۰٫۳	۹۳٫۲۹	۹۴٫۴۷	۹۴٫۳۴	۹۳٫۵۲	۹۴٫۲۱	۹۳٫۰۵	۹۵٫۸۸	پیشنهادی
	۰٫۶	۸۹٫۷۶	۹۱٫۴۲	۹۱٫۵۸	۸۹٫۵۳	۹۲٫۲۰	۹۱٫۹۴	۹۴٫۱۷	همبستگی منفی
	۰	۹۳٫۱۷	۹۳٫۷۶	۹۳٫۲۹	۹۳٫۰۶	۹۳٫۶۵	۹۲٫۹۴	۹۳٫۶۴	مستقل
۴	۰٫۴	۹۴٫۱۱		۹۴٫۴۷	۹۴		۹۳٫۰۵	۹۶٫۳۷	پیشنهادی
	۰٫۵	۹۳٫۳۲		۹۳٫۲۹	۹۲٫۳۵		۹۳٫۱۲	۹۴٫۱۸	همبستگی منفی
	۰	۹۴٫۰۵		۹۲٫۴۰	۹۳٫۶۵		۹۲٫۹۴	۹۳٫۶۳	مستقل
۲	۰٫۹		۹۲٫۷۶			۸۹٫۸۸		۹۵٫۷۱	پیشنهادی
	۰٫۸		۹۳٫۶۷			۹۲٫۹۴		۹۴٫۲۶	همبستگی منفی
	۰		۹۳٫۸۸			۹۲٫۱۸		۹۳٫۲۳	مستقل

۴-۲ آزمایش‌ها

برای نشان دادن کارایی روش پیشنهادی در ایجاد گوناگونی بین شبکه‌های پایه، آزمایش‌هایی انجام دادیم. این آزمایش‌ها برای حالتی که سیستم مرکب شامل دو، سه، چهار، پنج، شش، هشت و ده شبکه پایه باشد صورت گرفته است.

شبکه‌های عصبی استفاده شده به عنوان طبقه‌بندهای پایه، شبکه‌های پرسپترون با یک لایه ورودی، یک لایه مخفی و یک لایه خروجی هستند. تعداد نرون‌های لایه ورودی و خروجی به ترتیب برابر بعد بردار ویژگی و تعداد کلاس‌ها است. تعداد مناسب برای نرون‌های لایه مخفی، H، و تعداد دوره‌های[1] لازم، T، در مورد هر مجموعه داده با سعی و خطا بدست آمده است که در بخش بررسی نتایج تعداد آنها ذکر می‌شود.

آزمایش‌ها به ازای تغییر λ از ۰ تا ۰٫۹ با گام یکدهم انجام شد. وزنهای شبکه با الگوریتم یادگیری پس انتشار خطا و بر اساس تابع خطای (۱۴) تصحیح شدند.

نرخ یادگیری شبکه در هر دوره بر اساس رابطه $\eta = \eta_0(1-0.9t/T)$ بود که در آن T تعداد کل دوره‌ها، t شماره دوره جاری و η_0 نرخ یادگیری اولیه است. آزمایش‌ها به ازای مقادیر متنوعی از η_0 انجام شد ولی نتایج ارائه شده به ازای $\eta_0 = 0.9$ است.

برای آنکه تأثیر تصادفی بودن وزن‌های اولیه شبکه‌ها ناچیز شود، هر آزمایش را ۱۵ بار تکرار کردیم. بنابراین نتایج ارائه شده در این مقاله با متوسط‌گیری از نتایج ۱۵ بار اجرای مستقل هر آزمایش بدست آمده است. لازم به ذکر است که در همه آزمایش‌ها، نمونه‌های آموزشی و آزمایشی یکسان بوده‌اند، یعنی ابتدا نمونه‌ها را به دو گروه آموزشی و آزمایشی تقسیم کرده و هر آزمایش را بر روی این نمونه‌ها ۱۵ بار تکرار کرده‌ایم. به منظور مقایسه نتایج روش پیشنهادی با روشی مشابه، آزمایش‌ها را برای روش همبستگی منفی [۲۳] نیز تکرار کردیم. برای نشان دادن کارایی روش‌های مبتنی بر ایجاد گوناگونی، آزمایش‌ها را برای روش یادگیری مستقل نیز تکرار کردیم. در این روش یادگیری، هر شبکه به صورت مستقل از دیگر شبکه‌ها آموزش می‌بیند و سپس در مرحله آزمون، نتایج آنها با قاعده میانگین ترکیب شده و خروجی سیستم طبقه‌بندی مرکب را تشکیل می‌دهند. پیاده‌سازی این روش با انتخاب $\lambda = 0$ در (۱۴) تحقق می‌یابد.

1. Epochs

۴-۳ نتایج آزمایش‌ها و بررسی آنها

جدول ۲ بهترین نتایج در مورد هر روش و مقدار λ^* متناظر به آنرا بر روی مجموعه داده یونوسفر نشان می‌دهد. در این جدول میانگین نرخ بازشناسی شبکه‌های پایه و سیستم طبقه‌بندی مرکب آمده است. در تمام آزمایش‌ها از قاعده میانگین برای ترکیب نتایج خروجی طبقه‌بندها استفاده شده است. ستون‌های این جدول به ترتیب از راست به چپ تعداد شبکه‌های پایه، λی بهینه، میانگین نرخ بازشناسی شبکه‌های پایه، میانگین نرخ بازشناسی سیستم مرکب و روش یادگیری است.

در این جدول نرخ بازشناسی سیستم مرکب برای سه روش پیشنهادی، همبستگی منفی و یادگیری مستقل با یکدیگر مقایسه شده است. ملاحظه‌ی نرخ بازشناسی سیستم مرکب نشان می‌دهد که کارایی روش پیشنهادی ما از روش همبستگی منفی و روش یادگیری مستقل بیشتر است. علت این امر، توانمندتر بودن روش پیشنهادی برای ایجاد گوناگونی در شبکه‌های پایه است.

چنانکه جدول ۲، نشان می‌دهد، در روش یادگیری مستقل با ترکیب نتایج شبکه‌ها، بهبودی در نرخ بازشناسی سیستم مرکب حاصل نمی‌شود به‌طوری‌که نرخ بازشناسی سیستم مرکب از نرخ بازشناسی بهترین شبکه پایه کمتر است. علت این امر اشتراک زیاد بین خطاهای شبکه‌ها با یکدیگر است. ولی در روش همبستگی منفی و روش پیشنهادی ما، به علت گوناگونی در خطا، نرخ بازشناسی سیستم مرکب بیشتر از بزرگترین نرخ بازشناسی شبکه‌های پایه است. در جدول ۳ نتایج بعضی آزمایش‌های انجام‌شده بر روی مجموعه تصاویر ماهواره‌ای نشان داده شده است. چنانکه ملاحظه می‌شود، در مورد این مجموعه شش‌کلاسه نیز کارایی روش پیشنهادی ما نسبت به دو روش دیگر بهتر است. به عنوان نمونه برای حالتی که سیستم مرکب شامل پنج شبکه است، روش ما نسبت به روش‌های همبستگی منفی و مستقل به‌ترتیب ۱٫۲۳٪ و ۲٫۰۹٪ بهبود را بدست می‌دهد.

جدول ۴، نتایج برخی آزمایش‌های انجام‌شده را بر روی مجموعه دوکلاسه سونار نشان می‌دهد. در این حالت نیز در روش پیشنهادی و روش همبستگی منفی، به علت توانمندی برای ایجاد گوناگونی در شبکه‌های پایه، نرخ بازشناسی سیستم مرکب بیشتر از بزرگترین نرخ بازشناسی شبکه‌های پایه است. به عنوان مثال برای حالتی که تعداد اعضای سیستم مرکب شش شبکه است، روش پیشنهادی ما ۵٫۳۳٪ و روش همبستگی منفی ۴٫۹۹٪ بهبود نسبت به بهترین شبکه پایه ایجاد می‌کنند.

Archive of SID

www.SID.ir

جدول ۳: میانگین نرخ بازشناسی بر روی مجموعه تصاویر ماهواره‌ای: تعداد نرون‌های لایه مخفی هفت عدد، تعداد دوره‌ها در هر آزمایش ۲۰ و نتایج ارائه‌شده میانگین ۱۵ بار تکرار هر آزمایش است.

تعداد شبکه‌ها	λ^*	میانگین نرخ بازشناسی شبکه‌های پایه (٪)					میانگین نرخ بازشناسی سیستم مرکب (٪)	روش یادگیری
۵	۰٫۳	۸۵٫۱۹	۸۳٫۹۸	۸۵٫۵۸	۸۵٫۰۷	۸۵٫۶۳	۸۶٫۴۳	پیشنهادی
	۰٫۵	۸۴٫۳۸	۸۴٫۴۴	۸۴٫۴۲	۸۳٫۴۷	۸۴٫۴۱	۸۵٫۲۰	همبستگی منفی
	۰	۸۴٫۶۱	۸۴٫۸۴	۸۴٫۵۹	۸۴٫۵۱	۸۴٫۴۲	۸۴٫۳۴	مستقل
۳	۰٫۶	۸۱٫۳۴		۸۳٫۸۶		۸۵٫۶۴	۸۶٫۳۲	پیشنهادی
	۰٫۷	۸۴٫۲۱		۸۳٫۱۱		۸۲٫۴۷	۸۵٫۱۴	همبستگی منفی
	۰	۸۴٫۱۶		۸۴٫۳۸		۸۴٫۵۵	۸۴٫۵۴	مستقل
۲	۰٫۶	۸۵٫۵۴			۸۴٫۸۶		۸۶٫۱۸۳	پیشنهادی
	۰٫۲	۸۴٫۲۳			۸۴٫۵۳		۸۵٫۰۶	همبستگی منفی
	۰	۸۴٫۴۱			۸۴٫۴۸		۸۴٫۴۸	مستقل

جدول ۴: میانگین نرخ بازشناسی بر روی مجموعه داده سونار: تعداد نرون‌های لایه مخفی سه عدد، تعداد دوره‌ها در هر آزمایش ۴۰۰ و نتایج ارائه‌شده میانگین ۱۵ بار تکرار هر آزمایش است.

تعداد شبکه‌ها	λ^*	میانگین نرخ بازشناسی شبکه‌های پایه (٪)						میانگین نرخ بازشناسی سیستم مرکب (٪)	روش یادگیری
۶	۰٫۲	۸۰٫۷۴	۸۰٫۱	۸۱٫۱۹	۸۰٫۷۵	۸۱٫۹۴	۸۰٫۵۹	۸۶٫۴۸	پیشنهادی
	۰٫۸	۶۷٫۳۱	۶۹٫۴۰	۶۸٫۶۶	۶۷٫۱۶	۷۲٫۶۸	۷۱٫۹۴	۸۵٫۹۷	همبستگی منفی
	۰	۸۱٫۰۴	۸۱٫۷۸	۸۱٫۵۹	۸۲٫۰۹	۸۱٫۴۴	۸۱٫۶۶	۸۳٫۸۸	مستقل
۴	۰٫۴	۸۱٫۰۴		۷۷٫۷۶		۷۹٫۱۰	۸۱٫۲۴	۸۵٫۶۷	پیشنهادی
	۰٫۶	۷۹٫۸۵		۸۱٫۱۹		۷۹٫۷۰	۷۹٫۵۵	۸۵٫۴۶	همبستگی منفی
	۰	۸۱٫۷۵		۸۱٫۱۵		۸۲٫۲۴	۸۱٫۴۹	۸۳٫۴۳	مستقل
۲	۰٫۳	۸۲٫۸۴			۸۱٫۰۴		۸۵٫۳۸	پیشنهادی
	۰٫۶	۸۰٫۵۹			۸۱٫۴۹		۸۳٫۵۸	همبستگی منفی
	۰	۸۱٫۶۸			۸۱٫۷۵		۸۲٫۳۴	مستقل

جدول ۵: میانگین هندسی نسبت خطا

میانگین هندسی نسبت خطا	تعداد اعضای سیستم مرکب
۰٫۸۶	۲
۰٫۸۳	۴
۰٫۸۴	۶

چنانکه جدول‌های ۲ تا ۴ نشان می‌دهند، معمولاً مقادیر بزرگتر برای λی بهینه، بهبود بیشتری در نرخ بازشناسی سیستم مرکب، نسبت به بهترین شبکه پایه، بدست می‌دهند. علت این امر متفاوت‌تر شدن نمونه‌های خطای شبکه‌های پایه با یکدیگر است. به عنوان نمونه سطر دوم جدول ۴ متناظر با $\lambda=۰٫۸$ را ببینید. گرچه در این حالت نرخ بازشناسی شبکه‌های پایه کم است، ولی نرخ بازشناسی سیستم مرکب بهبود قابل توجهی نسبت به آنها دارد.

یکی از معیارهای آماری مناسب برای بررسی کارایی نسبی دو روش یادگیری استفاده از میانگین هندسی[1] نسبت خطاهای آنها روی مجموعه داده‌های متفاوت است [۲۷]. اگر E_p میانگین خطای روش پیشنهادی و E_n میانگین خطای روش همبستگی منفی روی یک مجموعه داده باشند، این معیار به صورت (۱۸) تعریف می‌شود

$$GMER=\sqrt[n]{\prod_{i=1}^{n}\frac{E_P}{E_N}} \qquad (۱۸)$$

که در آن n تعداد مجموعه داده‌هایی است که کارایی دو روش روی آنها ارزیابی شده است.

اگر نسبت E_P/E_N را با V نشان دهیم، معیار $GMER$ در مقاله ما به صورت زیر خواهد بود

$$GMER=\sqrt[3]{\prod_{i=1}^{3}\frac{E_P}{E_N}}=\sqrt[3]{\prod_{i=1}^{3}V_i}=\sqrt[3]{V_1.V_2.V_3} \qquad (۱۹)$$

که در آن V_1، V_2 و V_3 به ترتیب میانگین نسبت خطای دو روش بر روی مجموعه داده‌های یونوسفر، تصاویر ماهواره‌ای و سونار است. برای حالتی که سیستم مرکب شامل ۲، ۴ و ۶ عضو باشد، این معیار محاسبه گردیده و مقدار آن در جدول ۵ آمده است.

معمولاً مقدار کمتر از ۰٫۹ برای ارزیابی دو روش با یکدیگر مناسب است. چنانکه جدول ۵ نشان می‌دهد مقدار این معیار برای مقایسه کارایی روش پیشنهادی با روش همبستگی منفی مناسب است.

روش پیشنهادی بر روی مجموعه داده دوکلاسه فونم با ۳۶۰۲ نمونه آموزشی و ۱۸۰۲ نمونه آزمایشی نیز اعمال شد. کمترین و بیشترین اختلاف بین نرخ بازشناسی روش پیشنهادی با روش همبستگی منفی به ترتیب ۰٫۶۸٪، برای سیستم مرکب شامل ۶ عضو، و ۱٫۹۸٪ برای سیستم مرکب شامل ۳ عضو بود.

در تمام آزمایش‌های انجام شده از قاعده میانگین به عنوان قاعده ترکیب استفاده شده است. گرچه بکارگیری قاعده ترکیب دمپستر–شفر می‌تواند نتایج بهتری را حاصل کند، اما چون در این تحقیق هدف ما صرفاً مقایسه کارایی روش پیشنهادی با روش‌های یادگیری مستقل و

1. Geometric Mean Error Ratio

Archive of SID

www.SID.ir

همبستگی منفی بوده است، از متداولترین قاعده ترکیب استفاده کرده‌ایم.

۵- نتیجه‌گیری

در این تحقیق یک روش جدید یادگیری برای سیستم طبقه‌بندی مرکب پیشنهاد شد. این روش با ایجاد گوناگونی در خطا برای طبقه‌بندهای پایه، باعث افزایش کارایی طبقه‌بندی مرکب می‌شود. در این روش شباهت خطای هر طبقه‌بند با سایر طبقه‌بندها به عنوان یک مؤلفه در تابع خطای آن طبقه‌بند منظور می‌شود. هرچه شباهت خطای یک طبقه‌بند با سایر طبقه‌بندها بیشتر باشد، ضرایب یادگیری آن طبقه‌بند با شدت بیشتری تصحیح می‌شوند.

کارایی روش پیشنهادی، در حالتی که طبقه‌بندهای پایه از نوع شبکه‌های عصبی مصنوعی هستند، بر روی سه مجموعه داده بررسی شد. نتایج آزمایش‌ها، کارایی بهتر این روش را نسبت به روش مشابه آن، یادگیری با همبستگی منفی، نشان داد.

بررسی کارایی این روش نسبت به روشهای کیسه‌کردن و تقویتی که از پرکاربردترین روش‌های ضمنی و صریح در ایجاد گوناگونی هستند و همچنین تعیین محدوده مناسب برای λ و ارتباط آن با خصوصیات آماری مجموعه داده، از موضوعات مورد علاقه ما برای ادامه این تحقیق است.

مراجع

[1] V. Gunes and M. Menard, "Combination, cooperation and selection of classifiers: a state of the art," *Int. J. of Pattern Recognition and Artificial Intelligence*, vol. 17, no. 8, pp. 1303-1324, Dec. 2003.
[2] G. Rogova, "Combining the results of several neural network classifiers," *Neural Networks*, vol. 7, no. 5, pp. 777-781, 1994.
[3] 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.
[4] A. Krogh and J. Vedelsby, "Neural network ensembles, cross validation, and active learning," in *Proc. Advances in Neural Information Processing Systems*, vol. 7, pp. 231-238, 1995.
[5] S. Hashem, B. Schmeiser, and Y. Yih, "Optimal linear combinations of neural networks: an overview," in *IEEE Int. Conf. on Neural Networks*, vol. 3, pp. 1507-1512, 1994.
[6] R. Maclin and J. Shavlik, "Combining the predictions of multiple classifiers: using competitive learning to initialize neural networks," in *Proc. of 14th Int. Joint Conf. on Artificial Intelligence*, pp. 524-530, Montreal, Canada, Aug. 1995.
[7] 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.
[8] Y. Liu and X. Yao, "Ensemble learning via negative correlation," *Neural Networks*, vol. 12, no. 10, pp. 1399-1404, Dec. 1999.
[9] 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.
[10] W. Wang, P. Jones, and D. Partridge, "Diversity between neural networks and decision trees for building multiple classifier systems," in *Proc. Int. Workshop on Multiple Classifier Systems*, *Lecture Notes In Computer Science,* vol. 1857, pp. 240-249, Calgiari, Italy, 2000.
[11] R. P. W. Duin and D. M. J. Tax, "Experiments with classifier combining rules," in *Proc. Int. Workshop on Multiple Classifier Systems*, *Lecture Notes In Computer Science,* vol. 1857, pp. 16-29, Calgiari, Italy, 2000.
[12] M. Skurichina and R. P. W. Duin, "Bagging, boosting, and the random subspace method for linear classifiers," *Pattern Analysis and Applications*, vol. 5, no. 2, pp. 121-135, 2002.
[13] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp.123-140, 1996.
[14] Y. Raviv and N. Intrator, "Bootstrapping with noise: an effective regularization technique," *Connection Science*, vol. 8, no. 3-4, pp. 355-372, Dec. 1996.
[15] A. Sharkey, N. Sharkey, and G. Chandroth, "Diverse neural net solutions to a fault diagnosis problem," *Neural Computing and Applications*, vol. 4, no. 4, pp. 218-227, 1996.
[16] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proc. of the 13th Int. Conf. on Machine Learning*, pp. 148-156, 1996.
[17] Y. Liu and X. Yao, "Ensemble learning via negative correlation," *Neural Networks*, vol. 12, no. 10, pp. 1399-1404, 1999.
[۱۸] س. ح. نبوی کریزی و ا. کبیر، "ترکیب طبقه‌بندها: ایجاد گوناگونی و قواعد ترکیب"، مجله علوم و مهندسی کامپیوتر، مجلد۳، شماره ۳ (الف)، صص ۱۰۷-۹۵، پاییز ۱۳۸۴.
[19] N. Ueda and R. Nakano, "Statistical analysis of the generalization error of ensemble estimators," in *Int. Conf. on Neural Networks*, ICNN96, pp. 90-95, 1996.
[20] K. Tumer and J. Ghosh, "Analysis of decision boundaries in linearly combined neural classifiers," *Pattern Recognition*, vol. 29, no. 2, pp. 341-348, 1996.
[21] G. Brown, *Diversity in Neural Network Ensembles*, Ph.D. Thesis, University of Birmingham, Sep. 2003.
[22] B. E. Rosen, "Ensemble learning using decorrelated neural networks," *Connection Science*, vol. 8, no. 3-4, pp. 373-384, Dec. 1996.
[23] Y. Liu, *Negative Correlation Learning and Evolutionary Neural Network Ensembles*, Ph.D. Thesis, University of New South Wales, 1998.
[24] X. Yao, M. Fischer, and G. Brown, "Neural network ensembles and their application to traffic flow prediction in telecommunications networks," in *Proc. of Int. Joint Conf. on Neural Networks*, pp. 693-698, 2001.
[25] G. Brown and X. Yao, "On the effectiveness of negative correlation learning," in *Proc. First UK Workshop on Computational Intelligence*, pp. 57-62, Edinburgh, Scotland, Sep. 2001.
[26] www.dice.ucl.ac.be/neural-nets/Research/Projects/ELENA/database and www.ics.uci.edu/~mlearn.
[27] G. I. Webb, "Multi boosting: a technique for combining boosting and wagging," *Machine Learning*, vol. 40, no. 2, pp. 159-197, Aug. 2000.

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