افزایش نرخ کارایی طبقه‌بندی با استفاده از تجمیع ویژگی‌های مؤثّر روش‌های مختلف ترکیب شبکه‌های عصبی

جواد محمدزاده [۱]، سعید مسعودنیا [۲]، علی آرانی [۳]
[۱] دانشگاه آزاد اسلامی، واحد کرج، [۲] دانشگاه آزاد اسلامی، واحد تهران مرکز و باشگاه پژوهشگران جوان و
[۳] دانشگاه تربیت دبیر شهید رجایی

چکیده:

روش‌های یادگیری همبستگی منفی و اختلاط خبره‌ها، به‌عنوان دو روش معروف ترکیب شبکه‌های عصبی، از توابع خطای منحصر به فرد و البته متفاوتی برای آموزش شبکه‌های پایه، به‌صورت هم‌زمان استفاده کرده، که شبکه‌هایی با همگرایی منفی تولید می‌کنند. در این مقاله ویژگی‌های مختلف این دو روش را مرور و نقاط قوت و ضعف آنها را در مقایسه با یکدیگر بررسی می‌کنیم. بررسی خصوصیات این دو روش در مقایسه با یکدیگر نشان داد که آن‌ها ویژگی‌های متفاوت و البته مکملی نسبت به هم دارند؛ به نحوی که اگر بتوان سیستمی ترکیبی شامل ویژگی های مثبت هر دو روش طراحی کرد، به‌احتمال کارایی بهتری از روش‌های پایهٔ خود داشته باشد. در این مقاله ایده‌ای برای ترکیب ویژگی‌های این دو روش پیشنهاد کرده‌ایم. در این روش، قابلیت پارامتر کنترلی روش یادگیری همبستگی منفی به تابع خطای روش اختلاط خبره‌ها افزوده شده، که این روش را قادر می‌سازد تعادل بهینه‌ای را در توازن بایاس-واریانس-کوواریانس ایجاد و کارایی را افزایش دهد. روش ترکیبی پیشنهاد شده، در چند مسألهٔ آزمون پیش‌بینی و طبقه‌بندی با روش‌های پایهٔ اختلاط خبره‌ها و یادگیری همبستگی منفی مورد مقایسه قرار گرفته است. نتایج آزمایش‌ها نشان داده است که روش ترکیبی پیشنهاد شده، با حفظ نقاط قوّت و کاهش ضعف‌های روش‌های پایه، توانسته کارایی را به‌طور قابل ملاحظه‌ای نسبت به آنها افزایش دهد.

واژگان کلیدی: طبقه بندی، ترکیب شبکه‌های عصبی، یادگیری همبستگی منفی و اختلاط خبره‌ها.

۱- مقدّمه

ترکیب شبکه‌های عصبی(*Rokach, 2010b*) روشی برای بهبود کارایی در مسائل پیش‌بینی و طبقه‌بندی، به‌خصوص در حل مسائل پیچیده‌ای با وجود تعداد محدود نمونه، بردار ویژگی‌هایی با ابعاد بالا و یا دسته‌هایی با هم‌پوشانی زیاد است (*Ghaemi et al., 2010; Ebrahimpour et al., 2011; Rokach, 2010a*). روش‌های ترکیب از دو جزء اصلی تشکیل شده‌اند: ۱. قسمت آموزش شبکه‌های پایهٔ سیستم ترکیبی ۲. قسمت ترکیب خروجی‌ها. تحقیقات تئوری و نیز تجربی نشان می‌دهد که کارایی سیستم شبکه‌های عصبی، زمانی بهینه است که بین شبکه‌های پایه، همبستگی منفی وجود داشته باشد.روش‌های مختلفی برای ایجاد همبستگی منفی میان شبکه‌های پایه پیشنهاد شده است که از آن جمله می‌توان به وزن‌دهی اوّلیه به‌صورت تصادفی، استفاده از شبکه‌هایی با ساختارهای متفاوت،به‌کارگیری الگوریتم‌های متفاوت یادگیری و متفاوت ساختن مجموعه‌های یادگیری اشاره کرد. در مقالهٔ (*Islam et al., 2008b*) اشاره شده که متفاوت کردن مجموعه‌های یادگیری شبکه‌های پایه از بقیهٔ روش‌های مورد استفاده در تولید شبکه‌هایی با همبستگی منفی، مؤثّرتر عمل می‌کند.

دو روش ترکیب شبکه‌های عصبی که به‌ترتیب شبکه‌های پایه را به‌صورت مستقل از هم و پی‌درپی آموزش می‌دهند و از ایدهٔ ایجاد مجموعه‌های یادگیری متفاوت به‌صورت صریح استفاده می‌کنند، روش کیسه‌ای‌کردن[1] و تقویتی[2] نام دارند. روش‌های یادگیری همبستگی منفی[3] و اختلاط خبره‌های[4] پایه، به‌عنوان روش‌های دیگری در ترکیب طبقه‌بندها با به‌کارگیری توابع خطای ویژه‌ای،

[1] Bagging
[2] Boosting
[3] Negative Correlation Learning
[4] Mixture of Experts

Archive of SID
www.SID.ir

شبکه‌های پایه را به‌صورت هم‌زمان آموزش می‌دهند. درحالی‌که روش‌های کیسه‌ای‌کردن و تقویتی با تغییر توزیع احتمال انتخاب مجموعه یادگیری هر یک از خبره‌های پایه، صریحاً مجموعه‌های متفاوتی را برای شبکه‌های پایه لحاظ می‌کنند، روش‌های یادگیری همبستگی منفی و اختلاط خبره‌های پایه با به‌کارگیری توابع خطای ویژه‌ای، شبکه‌های پایه را به‌صورت ضمنی به یادگیری قسمت‌های متفاوتی از مجموعه یادگیری ترغیب می‌کنند(*Rokach, 2010b*). با توجه به نحوه ایجاد تفاوت در مجموعه‌های یادگیری، گروه اول را روش‌های صریح و گروه دوم را روش‌های ضمنی تقسیم فضای یادگیری می‌نامند. از آن‌جایی که این دو نوع روش، نقاط قوت و ضعف مختلف و در بعضی موارد مکملی دارند، تحقیقاتی سعی داشته‌اند تا ویژگی مکمل این روش‌ها را در یک سیستم ترکیبی جمع کنند، که مواردی از این تحقیقات در ادامه مرور خواهد شد.

واترهوس[1] در مقالهٔ (*Waterhouse, 1997*) ایده‌هایی برای ترکیب ویژگی‌های دو روش اختلاط خبره‌های پایه و روش تقویتی ارائه کرده است. او و همکارانش دو روش ارائه کردند تا نقاط ضعف هر دو روش را با به‌کارگیری ویژگی هایی از روش دیگری برطرف کنند. ایدهٔ اوّل به‌عنوان بهبودی بر روش اختلاط خبره‌ها قابل بررسی است. در این روش از ایدهٔ تقسیم فضای مبتنی بر روش تقویتی برای جهت‌دهی اوّلیهٔ شبکه‌های پایهٔ اختلاط خبره‌ها استفاده شده است. از آنجایی که روش تقویتی، فضای مسأله را بر اساس قابلیت شبکه‌های پایه در حلّ مسأله تقسیم‌بندی می‌کند، می‌توان این نحوهٔ تقسیم فضا را برای جهت‌دهی اوّلیهٔ شبکه‌های پایهٔ اختلاط خبره‌ها لحاظ کرد. زیرا ویژگی محلی‌سازی مورد نیاز آن روش را برآورده می‌کند. ایدهٔ دوم به‌عنوان بهبودی بر روش تقویتی ارائه شده که در آن، از روش پویای شبکهٔ میانجی برای ترکیب خروجی شبکه‌های پایهٔ روش تقویتی استفاده شده است.

آونیملچ[2] و اینراتر[3] (*Avnimelech and Intrator, 1999*) ایده‌های کار واترهوس (*Waterhouse, 1997*) را گسترش و یک سیستم پویای ترکیب شبکه‌های عصبی را برای تجمیع ویژگی‌های دو روش اختلاط خبره‌های پایه و روش تقویتی پیشنهاد کردند. در مقالهٔ آنها ابتدا نقاط قوّت و ضعف این دو روش در مقایسه با یکدیگر بررسی شده است. در جهت بهبود روش‌های قبلی، آنها سیستم پویایی از ترکیب شبکه‌های عصبی ارائه کردند که از شبکهٔ میانجی مبتنی بر معیار اطمینان برای ترکیب خروجی‌ها استفاده می‌کند. روش پیشنهادی آنها توانست روش قبلی را پشت سر گذاشته و کارایی بالاتری نسبت به آنها به‌دست آورد.

ایسلام[4] و همکارانش در مقالهٔ (*Islam et al., 2008a*)، روش یادگیری همبستگی منفی را در الگوریتم روش‌های تقویتی و کیسه‌کردن وارد کردند و بدین‌وسیله کارایی آنها را بهبود دادند. ایدهٔ این روش در ساختار پویا و قابل گسترش ترکیب شبکه‌های عصبی پیاده‌سازی شده، به‌طوری که علاوه‌بر روش یادگیری متداول و جدای از هم شبکه‌های پایه در دو روش تقویتی و کیسه‌کردن، بازه‌ای از فرآیند یادگیری شبکه‌های پایه نیز به یادگیری هم زمان مبتنی بر همبستگی منفی اختصاص داده شده است.

در ادامه، ویژگی‌های دو روش یادگیری همبستگی منفی و اختلاط خبره‌های پایه را بررسی و نقاط ضعف و قوّت آنها را در مقایسه با یکدیگر مشخّص کرده‌ایم. از آنجایی که نقاط قوّت و ضعف این دو روش مکمل یکدیگر به‌نظر می‌رسند، سؤالی مطرح می‌شود که آیا جمع‌کردن ویژگی‌های این دو روش در ساختار یک روش ترکیبی مجتمع می‌تواند موجب بهبود کارایی شود یا خیر؟ با بررسی این سؤال، ایده‌ای برای ترکیب ویژگی‌های روش‌های اختلاط خبره‌های پایه و یادگیری همبستگی منفی ارائه کرده‌ایم. در روش پیشنهادی با استفاده از یادگیری همبستگی منفی مقادیر پارامترهای منظم‌سازی در روش اختلاط خبره‌ها تنظیم و به این ترتیب کارایی این روش بهبود یافته است.

۲- مروری بر روش‌های اختلاط خبره‌ها و یادگیری همبستگی منفی

در این بخش، روش‌های اختلاط خبره‌های پایه و یادگیری همبستگی منفی و ویژگی‌های این دو روش ترکیب شبکه‌های عصبی مرور می‌شود.

۲-۱- اختلاط خبره‌های پایه

مدل اختلاط خبره‌ها توسط جیکابز[5] و همکارانش در مقالات (*Jacobs et al., 1991*) معرفی شد. در این مقالات توابع خطای مختلف برای استفاده به‌عنوان تابع خطای ترکیب شبکه‌های عصبی مورد آزمون قرار گرفته است. در این مقاله اشاره شده که محلی‌سازی خبره‌ها در نواحی مختلفی از

[1] Waterhouse
[2] Avnimelech
[3] Intrator
[4] Islam
[5] Jacobs

www.SID.ir
Archive of SID

فضای مسأله و در نتیجهٔ افزایش گوناگونی میان آنها موجب بهبود کارایی سیستم ترکیبی خواهد شد. بر مبنای این معیارها، در این مقاله جیکابز تابع خطای جدید برای ترکیب شبکه‌های عصبی معرفی می‌کند که بر مبنای احتمال تولید بُردار خروجی مطلوب در مدل ترکیبی از گوسین‌ها کار می‌کند:

$$E = -\log \sum_j g_j \ \exp\left(-\frac{1}{2}\left(y - O_j\right)^2\right) \qquad (۱)$$

برای بررسی نحوهٔ عملکرد این تابع خطا، مشتق آن نسبت به خبرهٔ iاُم بررسی شده است

$$\frac{\delta E}{\delta O_i} = -\left[\frac{g_i \ \exp\left(-\frac{1}{2}(y - O_i)^2\right)}{\sum_j g_j \ \exp\left(-\frac{1}{2}(y - O_j)^2\right)}\right](y - O_i) \qquad (۲)$$

بر اساس مشتق تابع خطای سوم، اصلاح وزن هر خبره، مشابه تابع خطای دوم، بر مبنای خطای مجزای آن صورت می‌گیرد. امتیاز این تابع خطا نسبت به موارد قبل وجود ضریبی برای اصلاح وزن هر خبره است که مقدار این ضریب با کارایی نسبی آن خبره نیز متناسب است. این دو ویژگی تابع خطای سوم در کنار هم موجب محلی‌سازی هر خبره در محدوده زیر فضای اختصاصی خودش می‌شود. این تابع خطا با برطرف‌کردن نواقص توابع خطای اوّل و دوم موجب کارایی بهتری در مدل اختلاط خبره‌های پایه می‌شود *(Jacobs et al., 1991)*.

همچنین، در این مدل ترکیبی یک شبکهٔ میانجی برای مدل‌سازی نحوهٔ پخش و تقسیم فضای خبره‌ها به‌کار گرفته شده است. شبکهٔ میانجی در حین فرآیند یادگیری خود، میزان خبرگی هر شبکهٔ عصبی در نواحی مختلف مسأله را آموزش دیده و نیز با دادن ضرایبی به میزان اصلاح وزن هر خبره، در نحوهٔ محلی‌سازی آن دخالت می‌کند. قانون یادگیری و اصلاح وزن شبکهٔ میانجی به‌نحوی است که با فرض مدل اختلاط گوسین‌ها، فرآیند یادگیری، میزان درست‌نمایی فضای مسأله را در این مدل افزایش می‌دهد؛ به‌صورتی که هر خبره مسئول یکی از این اجزای ترکیب گوسین‌ها باشد. به این ترتیب شبکهٔ میانجی فضای مسأله را تقسیم و هر خبره را به‌سمت زیر فضای لحاظ شده برای آن سوق می‌دهد که این امر بر مبنای نظر شبکهٔ میانجی با لحاظ‌کردن کارایی هر خبره صورت می‌گیرد.

مدل اختلاط خبره‌های پایه، ویژگی‌های جالبی دارد که آن را از سایر روش‌های ترکیب مجزا می‌کند. یکی از تمایزات این روش، ایدهٔ ترکیب پویای وزن‌دار خبره‌ها بر مبنای خبرگی آنها در زیر فضاهای مختلف مسأله است. در این مدل بر مبنای اصل تقسیم و غلبه، فضای پیچیده مسأله به تعدادی زیر فضای ساده‌تر بین خبره‌ها تقسیم می‌شود. در این مدل، بر خلاف روش‌های دیگر ترکیب شبکه‌های عصبی، به جای نسبت‌دادن وزن‌های ثابت برای ترکیب خبره‌ها، یک شبکهٔ میانجی به سیستم ترکیب اضافه شده است که بر حسب ورودی و خبرگی محلی خبره‌ها در نواحی مختلف فضای مسأله، وزن‌های بهینهٔ ترکیب شبکه‌های پایه را تخمین بزند *(Masoudnia and Ebrahimpour, 2012)*. به‌دلیل این ویژگی‌های مهم، این روش در تحقیقات مختلف یادگیری ماشین *(Masoudnia et al., 2010; Ubeyli, 2009)* و نیز در روش‌های پردازش الگو *(Ebrahimpour et al., 2011; Ubeyli et al., 2010)* بسیار مورد توجه قرار گرفته است.

۲-۲- یادگیری همبستگی منفی

به‌طور معمول در روش‌های مختلف ترکیب شبکه‌های عصبی، شبکه‌های پایه به‌صورت مستقل از هم مورد آموزش قرار می‌گیرند. یکی از معایب این روش‌ها فقدان ارتباط شبکه‌های پایه و همکاری میان آنها در حین فرآیند یادگیری است. یادگیری مستقل شبکه‌های پایه سیستم ترکیبی موجب می‌شود تا آنها نتوانند در حلّ مسأله به‌صورت مؤثّری با هم همکاری کنند.

لیو و یاو[1] در مقاله‌های *(Liu and Yao, 1999a, b)* روش یادگیری همبستگی منفی را معرّفی کردند که در آن شبکه‌های عصبی به‌صورت هم‌زمان و با همکاری یکدیگر به حلّ مسأله می‌پردازند. این همکاری و ارتباط چندجانبه با لحاظ‌کردن عبارت خطای همبستگی در تابع خطای هر یک از آنها به‌دست آمده است. در این روش، تابع خطای شبکهٔ عصبی iاُم ترکیب به این ترتیب است:

$$E_i = \frac{1}{2}\left(O_i - y\right)^2 + \lambda P_i \qquad (۳)$$

که O_i و y به‌ترتیب خروجی واقعی شبکهٔ iاُم و خروجی مطلوب است. اولین عبارت در تابع خطای مذکور،

[1] Liu and Yao

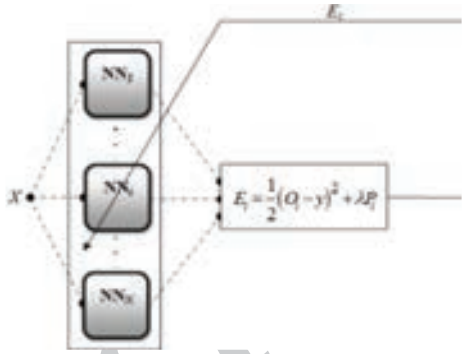
Archive of SID
www.SID.ir

تابع ریسک تجربی شبکهٔ iاُم است. عبارت دوم P_i عبارت جریمهٔ همبستگی است، که طبق رابطهٔ ذیل محاسبه می‌شود.

$$P_i = -\left(O_i - \overline{O}\right)^2 \quad (۴)$$

که $\overline{O}$ میانگین خروجی شبکه‌های عصبی پایه است.

طرح کلّی روش یادگیری همبستگی منفی بر اساس تابع خطای مذکور در (شکل ۱) نمایش داده شده است.



(شکل ۱): طرح کلی روش یادگیری همبستگی منفی. در مرحلهٔ یادگیری این روش، همان‌طور که در شکل نمایش داده شده است، شبکه‌های پایه با استفاده از تابع خطای ویژه یادگیری همبستگی منفی به‌طور هم‌زمان مورد آموزش قرار می‌گیرند.

عبارت P_i را می‌توان به‌عنوان یک عبارت منظم‌سازی[1] تعبیر کرد که در تابع خطای هر شبکهٔ ترکیب لحاظ شده است. به‌کارگیری این قبیل عبارت‌های منظم‌سازی راه حل مرسومی برای ایجاد تعادل بهینه در توازن بایاس‌ـواریانس‌ـکوواریانس[2] محسوب می‌شود. این عبارت برای محاسبهٔ مقدار همبستگی خروجی‌ها به‌کار گرفته شده که در فرآیند همگرایی، کمینه شده و منجر به شبکه‌هایی با همبستگی منفی شود. ضریب λ پارامتری برای تنظیم میزان تأثیر عبارت همبستگی در مقابل تابع ریسک تجربی است که با تغییر آن می‌توان به تعادل بهینه بین تابع خطای هدف و عبارت جریمهٔ همبستگی نزدیک شد. در هنگامی که این ضریب صفر لحاظ شود، عبارت همبستگی حذف شده و شبکه‌های سیستم ترکیبی به صورت مستقل از یکدیگر آموزش می‌بینند. میزان ارتباط و همبستگی میان شبکه‌های پایه با مقدار ضریب λ به‌صورت صریح قابل کنترل است. این تابع جریمه، شبکه‌های پایهٔ سیستم ترکیبی را تشویق می‌کند تا هر یک قسمت‌ها یا جنبه‌های مختلفی از مسأله را یاد گرفته و مدل کنند؛ تا در همکاری آنها با یکدیگر، سیستم ترکیبی بتواند کلّ فضای مسأله را به‌خوبی پوشش دهد *(Liu and Yao, 1999b)*.

براون در مقالهٔ *(Brown and Wyatt, 2003)* نشان داد که روش یادگیری همبستگی منفی را می‌توان مشتق شده از نحوهٔ خاصی از تجزیهٔ ابهام[3] به حساب آورد. تجزیه ابهام، به‌عنوان یکی از مهم‌ترین نتایج تئوری مبحث ترکیب یادگیرنده‌ها مطرح است. از این تجزیه، قابل اثبات است که میانگین مربع خطای ترکیب از میانگین مربع خطای شبکه‌های پایهٔ ترکیب، به اطمینان کمتر است.

$$(O_{ens} - y)^2 = \sum_j g_j (O_j - y)^2 - \sum_j g_j (O_j - O_{ens})^2 \quad (۵)$$

که در این رابطه g ضریب شبکه iام و O_{ens} ترکیب محدب خروجی شبکه‌های پایه است. تجزیه ابهام توجیه واضح و شفافی برای نحوهٔ تأثیر میزان همبستگی بر کارایی ترکیب ارائه می‌کند. این تجزیه، خطای کلّی ترکیب را به دو عبارت مستقل تجزیه می‌کند. عبارت اوّل میانگین وزن‌دار خطای شبکه‌های پایه سیستم ترکیبی است. عبارت دوم که عبارت ابهام نامیده شده، میزان گوناگونی شبکه‌های پایهٔ سیستم ترکیبی را اندازه‌گیری می‌کند. همان‌طور که در قبل هم اشاره شد، به‌کارگیری ضریب λ این امکان را می‌دهد تا میزان تأثیر عبارت همبستگی در مقایسه با عبارت خطای یادگیرنده‌ها کنترل شود. تحت کنترل درآوردن میزان این تأثیر، منجر به ایجاد تعادل کمینه در توازن بین این دو عبارت می‌شود. توازن بین این دو عبارت در ادبیات موضوع ترکیب یادگیرنده‌ها، در اصطلاح، توازن بین دقّت و گوناگونی[4] نامیده شده است.

در مقالهٔ *(Hansen, 2000)* ارتباط میان دو تجزیهٔ مختلف خطا، یعنی تجزیهٔ بایاس‌ـ واریانس‌ـکوواریانس و تجزیهٔ ابهام مورد بررسی قرار گرفته و نشان داده شده است که بخش‌های مختلف تجزیهٔ اوّل در تناسب با بخش‌های تجزیهٔ دوم قرار دارند. این ارتباط بین دو تجزیه، بیان‌گر آن است که ایجاد تعادل کمینه در توازن دقّت‌ـگوناگونی به ایجاد تعادل معادل در توازن بایاس‌ـواریانس-کوواریانس منجر خواهد شد و برعکس، افزایش قابلیت تعمیم روش، نتیجهٔ اصلی توازن این پارامترهاست.

۳- روش ارائه شده: اختلاط خبره‌هایی با همبستگی منفی

در این بخش، در ابتدا ویژگی‌های روش اختلاط خبره‌های پایه در مقایسه با روش یادگیری همبستگی منفی

[1] Regularization Term
[2] Bias-Variance-Covariance Tradeoff
[3] Ambiguity
[4] Accuracy-Diversity Tradeoff

www.SPA.ir
Archive of SID

بررسی و نقاط ضعف و قوّت آنها در مقایسه با یکدیگر بررسی می‌شود. بر اساس ساختار و راهبرد یادگیری مشابهی که در هر دو روش اختلاط خبره‌های پایه و یادگیری همبستگی منفی وجود دارد و بر مبنای ویژگی‌های مکمل آن‌ها، ایدهٔ ترکیب ویژگی‌های این دو روش در قالب یک سیستم ترکیب مجتمع ممکن است بتواند با حفظ نقاط قوت روش‌های پایه و حذف نقاط ضعف‌شان کارایی را افزایش دهد. بر مبنای این ایده، در این بخش، روش ترکیبی جدید اضافه‌کردن یادگیری همبستگی منفی به الگوریتم همگرایی روش اختلاط خبره‌های پایه برای جمع‌کردن ویژگی‌های روش اختلاط خبره‌های پایه و یادگیری همبستگی منفی پیشنهاد شده است:

۳-۱- مقایسهٔ نقاط ضعف و قوّت

در این بخش با بررسی ویژگی‌های روش‌های اختلاط خبره‌های پایه و یادگیری همبستگی منفی، به مقایسهٔ نقاط ضعف و قوت آن‌ها در مقایسه با یکدیگر خواهیم پرداخت. در ابتدا، ویژگی‌های مشترک این دو روش بررسی خواهد شد. هر دو روش ترکیب، شبکه‌های عصبی پایه را به صورت هم‌زمان و در فرآیندی با ارتباط و همکاری چندجانبه میان خبره‌ها آموزش می‌دهند. همان‌طور که در قبل هم ذکر شد، توابع متفاوت و ویژهٔ این دو روش، ویژگی‌هایی دارند که موجب می‌شود خبره‌ها در فرآیندی رقابتی توأم با همکاری، زیرفضاها یا جنبه‌های مختلفی از فضای مسأله را آموزش دیده، به‌نحوی که سیستم ترکیبی درمجموع بتواند کلّ فضای مسأله را مدل‌سازی کند. به عبارت دیگر، توابع خطای این دو روش ویژگی‌هایی دارند که موجب می‌شود با تقسیم فضای ضمنی مسأله بین خبره‌ها، شبکه‌های عصبی پایه‌ای تولید کنند که علاوه بر بایاس بودن، دارای همبستگی منفی نیز باشند *(Islam et al., 2008a)*.

در کنار این شباهت‌ها، این دو روش تفاوت‌هایی نیز با هم دارند. یکی از برتری‌های اختلاط خبره‌های پایه بر روش‌های ترکیبی مشابه دیگر، روش منحصر به فرد آن در ترکیب خروجی خبره‌هاست. اختلاط خبره‌های پایه از روشی مبتنی بر یادگیری برای ترکیب خروجی خبره‌ها استفاده می‌کند، که این قابلیت را دارد تا با مدل‌سازی میزان خبرگی شبکه‌های پایه در زیرفضاهای مختلف مسأله، به‌صورت پویا و بر حسب ورودی، خروجی خبره‌ها را به‌صورت وزن‌دار ترکیب کند. شبکه‌ی میانجی در مدل اختلاط خبره‌های پایه، خروجی خبره‌ها را طبق میانگین وزن‌دار ترکیب می‌کند، به‌نحوی که وزن‌های ترکیب، تابعی از میزان خبرگی هر خبره به‌ازای هر ورودی خاص باشد.

همان‌طور که در قبل ذکر شد، سیستم‌های ترکیب از دو جزء اصلی تشکیل شده‌اند. در جزء اوّل، یعنی آموزش شبکه‌های عصبی پایه، بر طبق برتری خواص ذکر شده در فصل قبل(دارا بودن امکانی برای تنظیم پارامتر منظم‌سازی)، یادگیری همبستگی منفی به‌نحو کاراتری از اختلاط خبره‌های پایه رفتار می‌کند. برتری یادگیری همبستگی منفی از آنجا ناشی می‌شود که عبارت منظم‌سازی به‌کار گرفته شده در این روش موجب رسیدن این سیستم ترکیبی به تعادل بهینه در توازن بایاس-واریانس-کوواریانس و درنتیجه قابلیت تعمیم بهتر می‌شود؛ درحالی‌که اختلاط خبره‌های پایه فاقد این عبارت منظم‌سازی است *(Liu and Yao, 1999a)*. از طرف دیگر، اختلاط خبره‌های پایه از نظر جزء دوم ترکیب، دارای روشی کاراتر برای ترکیب خروجی خبره‌ها می‌باشد*(Kuncheva, 2004)*.

همان‌طور که از بررسی ویژگی‌های دو روش اختلاط خبره‌های پایه و یادگیری همبستگی منفی و بررسی نقاط ضعف و قوت هر یک در مقایسه با یکدیگر مشخّص شد، این دو، ویژگی‌های مکملی دارند. بر مبنای این ایده به‌نظر می‌رسد که اگر بتوانیم سیستم مجتمعی طراحی کنیم که شامل ویژگی‌های مثبت هر دو روش باشد، این سیستم کارایی بهتری نسبت به اجزای تشکیل دهندهٔ خود داشته باشد. در بخش بعدی، بر مبنای ایده‌ی تجمیع ویژگی، روش ترکیبی بهبود داده شده‌ای ارائه شده است که سعی در ترکیب ویژگی‌های مثبت روش‌های پایه‌ی اختلاط خبره‌ها و یادگیری همبستگی منفی دارد.

۳-۲- افزودن یادگیری همبستگی منفی به روش اختلاط خبره‌ها

روش‌های اختلاط خبره‌های پایه و یادگیری همبستگی منفی از توابع خطای مختلفی برای یادگیری و ایجاد همبستگی منفی میان شبکه‌های پایه استفاده می‌کنند. اگرچه روش اختلاط خبره‌های پایه نیز خبره‌هایی با همبستگی منفی تولید می‌کند، امّا مانند روش یادگیری همبستگی منفی دارای پارامتر کنترلی برای تنظیم صریح و بهینهٔ توازن بایاس-واریانس-کوواریانس نیست. برای این‌که این امکان یادگیری همبستگی منفی را به روش اختلاط خبره‌های پایه بیفزاییم، پارامتر کنترلی یادگیری همبستگی

Archive of SID
www.SID.ir

منفی را به تابع خطای اختلاط خبره‌های پایه اضافه کرده‌ایم. پارامتر کنترلی P_j که به‌عنوان عبارت منظم‌سازی به تابع خطای اختلاط خبره‌های پایه افزوده شده و تابع خطای اصلاح شده مدل اختلاط خبره‌ها در روابط ذیل آورده شده است.

$$E = -\log \sum_{j=1}^{N} g_j \left(-\frac{1}{2}\left(y - O_j\right)^2 + \lambda P_j \right) \qquad (۶)$$

$$P_j = -\left(O_j - \overline{O}\right)^2 \qquad (۷)$$

عبارت کنترلی اضافه شده، همانند روش یادگیری همبستگی منفی، این امکان را فراهم می‌کند تا تعادل شبه بهینه در توازن بایاس-واریانس-کوواریانس ایجاد و درنتیجه قابلیت تعمیم روش اختلاط خبره‌های پایه بهبود پیدا کند. به‌دلیل این ویژگی، روش بهبودیافته اختلاط خبره‌های ارائه شده را به نام روش اختلاط خبره‌های با همبستگی منفی[1] نام‌گذاری کرده‌ایم.

در ساختار روش اختلاط خبره‌های با همبستگی منفی، هر خبره یک شبکهٔ پرسپترون چندلایه با یک لایهٔ مخفی در نظر گرفته شده است. با لحاظ کردن تابع خطای جدید، قوانین اصلاح وزن خبره‌ها در فرآیند همگرایی با استفاده از الگوریتم پس‌انتشار خطا[2] در روابط ذیل آمده است:

$$h_{MNCE,\,i} = -\left[\frac{g_i \exp\left(-\frac{1}{2}(y - O_i)^2 + \lambda P_i \right)}{\sum_{j=1}^{N} g_j \exp\left(-\frac{1}{2}(y - O_j)^2 + \lambda P_j \right)} \right] \qquad (۸)$$

$$\Delta w_{\,h,i} = \eta_e h_{MNCE,i} w_y^{\,T} \left[\left(y - O_i\right) - \lambda \frac{\partial P_i}{\partial O_i} \right] \qquad (۹)$$

$$(O_i(1-O_i))O_{h_i}(1-O_{h_i})x_i$$

$$\Delta w_{\,y,i} = \eta_e h_{MNCE,i} \left[\left(y - O_i\right) - \lambda \frac{\partial P_i}{\partial O_i} \right] \qquad (۱۰)$$

$$(O_i(1-O_i))O_{hi}^{T}$$

$$\frac{\partial P_i}{\partial O_i} = -2\left(1 - \frac{1}{N}\right)\left(O_i - \overline{O}\right) \qquad (۱۱)$$

که η_e نرخ یادگیری، λ ضریب پارامتر کنترلی، g_i iاُمین خروجی شبکهٔ میانجی بعد از اعمال عملگر بیشینهٔ نُرم، W_h و W_y به‌ترتیب وزن‌های لایهٔ پنهان و لایهٔ خروجی شبکه‌های خبره است. O_h^T ترانهاده بُردار O_h، خروجی لایهٔ مخفی شبکه‌های پایه است.

شبکهٔ میانجی در روش بهبودیافته به‌طور تقریبی مشابه روش قبلی عمل می‌کند؛ با این تفاوت که تابع خطای شبکهٔ میانجی، با توجّه به بهبود تابع خطای کلّ سیستم ترکیب، طبق رابطهٔ زیر محاسبه می‌شود:

$$E_{G,MNCE} = \frac{1}{2}(h_{MNCE} - O_g)^2 \qquad (۱۲)$$

همان‌طورکه در رابطه مشاهده می‌شود، پارامتر h آن با روابط قبلی شبکهٔ میانجی متفاوت است؛ زیرا با اصلاح تابع خطای سیستم ترکیب، پارامتر h نیز تغییر کرده است. بدین ترتیب قوانین یادگیری شبکهٔ میانجی نیز مشابه قبل طبق روابط ذیل محاسبه می‌شود:

$$\Delta w_{yg} = \eta_g (h_{MNCE} - O_g)(O_g(1-O_g))O_{hg}^{T} \qquad (۱۳)$$

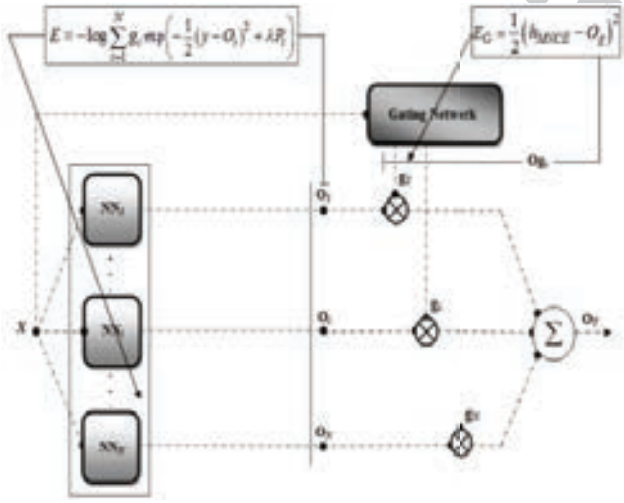
$$\Delta w_{hg} = \eta_g w_{yg}^{\,T} (h_{MNCE} - O_g) \qquad (۱۴)$$

$$(O_g(1-O_g))O_{hg}(1-O_{hg})x_i$$

خروجی نهایی سیستم نیز مشابه اختلاط خبره‌ها به‌صورت میانگین وزن‌دار طبق رابطهٔ ذیل محاسبه می‌شود:

$$O_T = \sum_{j=1}^{N} O_j g_j \qquad (۱۵)$$

ساختار روش اختلاط خبره‌هایی با همگرایی منفی و الگوریتم یادگیری هم‌زمان خبره‌ها و شبکهٔ میانجی در (شکل)۲ نمایش داده شده است.



(شکل۲): شمای کلّی الگوریتم یادگیری اختلاط خبره‌هایی با همبستگی منفی. شمای کلّی نحوهٔ یادگیری هم‌زمان شبکه‌های خبره و شبکهٔ میانجی با استفاده از توابع خطای اصلاح شده در شکل نمایش داده شده است. شبکه‌های خبره در یک فرآیند رقابتی با همکاری شبکهٔ میانجی به تقسیم فضا و مدل‌سازی حلّ مسأله می‌پردازند. پارامتر کنترلی اضافه شده λP_i این امکان را فراهم می‌کند تا میزان همبستگی منفی میان شبکه‌های خبره به‌صورت شبه بهینه تنظیم شود.

[1] Mixture of Negatively Correlated Experts
[2] Back-Propagation

www.SID.ir

بر پایهٔ الگوریتم یادگیری روش اختلاط خبره‌های با همبستگی منفی، خبره‌های سیستم ترکیبی در یک فرآیند یادگیری رقابتی به تقسیم فضای مسأله و مدل‌سازی آن می‌پردازند. شبکهٔ میانجی با مدل‌سازی میزان خبرگی محلّی شبکه‌های خبره، با تنظیم ضریب یادگیری آنها، این فرآیند تقسیم فضای میان خبره‌ها را مدیریت می‌کند. این فرآیند یادگیری رقابتی منجر به محلّی‌سازی خبره‌ها در زیرفضاهای خود، که به‌احتمال هم‌پوشانی نیز دارند، می‌شود. همچنین لحاظ‌کردن پارامتر کنترلی همبستگی منفی در تابع خطا به سیستم کمک می‌کند تا مقدار همبستگی منفی میان خبره‌ها را به نحو مطلوب تنظیم کند. هم‌چنین لحاظ‌کردن پارامتر کنترل همبستگی منفی در تابع خطا نیز به سیستم کمک می‌کند تا مقدار همبستگی منفی میان خبره‌ها را به نحو مطلوب تنظیم کند. همان‌طور که قبلاً توضیح داده شد، تنظیم مناسب میزان همبستگی میان خبره‌ها منجر به ایجاد تعادل شبه‌بهینه در توازن بایاس-واریانس-کوواریانس شده که این امر قابلیت تعمیم سیستم ترکیبی به ازای داده‌های آزمون را افزایش می‌دهد.

۴- نتایج آزمایش‌ها

در این بخش، آزمایش‌هایی برای بررسی کارایی روش ارائه شده، انجام شده است. در آزمایش اوّل، وابستگی کارایی روش اختلاط خبره‌هایی با همگرایی منفی به پارامتر کنترلی λP_i بررسی شده است. آزمایش دوم برای مقایسهٔ کارایی روش اختلاط خبره‌هایی با همگرایی منفی در مقابل روش‌های ترکیب پایهٔ اختلاط خبره‌ها و یادگیری همبستگی منفی انجام شده است. برای اندازه‌گیری کارایی روش‌ها، چندین مسألهٔ واقعی طبقه‌بندی از مجموعه داده‌های آزمون UCI به کار گرفته شده است. ویژگی‌های مجموعه داده‌های آزمون طبقه‌بندی مورد استفاده در (جدول ۱) به‌طور خلاصه ذکر شده است.

جدول ۱): اطّلاعات اصلی مجموعه داده‌های آزمون معیار طبقه‌بندی

مجموعه‌داده	تعداد نمونه	بعد بردار ویژگی	تعداد دسته
آواها[1]	۵۴۰۴	۵	۲
تصاویر ماهواره‌ای[2]	۶۴۳۵	۴	۶
ارقام دست‌نویس[3]	۱۰۹۹۲	۱۶	۱۰
شیشه[4]	۲۱۴	۱۰	۶
وسایل نقلیه[5]	۹۴۶	۱۸	۴

[1] Phoneme
[2] Satellite Images
[3] Pen Digit
[4] Glass
[5] Vehicle

در آزمایش‌های قبلی، روش‌های مورد بررسی در حل مسائل طبقه‌بندی مورد آزمون و مقایسه قرار می‌گرفتند؛ درحالی‌که در آزمایش سوم فرآیند همگرایی و یادگیری روش اختلاط خبره‌هایی با همگرایی منفی و اختلاط خبره‌های پایه را در مسئله رگرسیون و پیش‌بینی مورد بررسی قرار داده‌ایم. در این آزمایش، فرآیند یادگیری اختلاط خبره‌های پایه و اختلاط خبره‌هایی با همگرایی منفی از نقطه‌نظر نحوهٔ توازن عبارت‌های بایاس-واریانس-کوواریانس بررسی شده است. آزمایش آخر بدین منظور طراحی شده است تا علّت برتری اختلاط خبره‌هایی با همگرایی منفی بر اختلاط خبره‌های پایه را از لحاظ معیارهای یادگیری ماشین بررسی کند.

۴-۱- پارامترهای آزمون

در تمام آزمایش‌ها به‌جز آزمایش دوم، همهٔ سیستم‌های ترکیبی از پنج شبکه پرسپترون چندلایه با یک لایهٔ مخفی تشکیل شده‌اند. در آزمایش دوم، برای بررسی نحوهٔ تأثیر اندازهٔ سیستم ترکیبی بر کارایی روش‌های مورد مقایسه، سیستم‌های ترکیبی با اندازه‌های متنوعی شامل سه، پنج و هفت شبکهٔ پایه آزموده شده‌اند. تعداد نورون‌ها در لایهٔ مخفی برای تمام آزمایش‌ها، پنج نورون در نظر گرفته و الگوریتم یادگیری همهٔ روش‌های ترکیبی بر اساس روش پس‌انتشار خطا کار می‌کند. در روش‌های اختلاط خبره‌های پایه، اختلاط خبره‌هایی با همگرایی منفی و ترکیب شبکه‌های روش یادگیری همبستگی منفی با شبکهٔ میانجی، خبره‌ها و شبکهٔ میانجی به ترتیب با مقادیر نرخ یادگیری η_e=0.1 و η_g =0.05 مورد آموزش قرار می‌گیرند. نرخ یادگیری شبکه‌های پایهٔ روش یادگیری همبستگی منفی نیز مقدار 0.1 در نظر گرفته شده است. برای روش‌های یادگیری همبستگی منفی و اختلاط خبره‌هایی با همگرایی منفی، λ^* به‌عنوان مقدار بهینه λ، برای کم‌ترین نرخ خطا در یک فرآیند آزمون و خطا با آزمودن مقادیر مختلف بازه [۰:۰.۱:۲] (از صفر تا دو با گام‌های یک دهم) به‌ازای مجموعهٔ اعتبارسنجی[6] محاسبه و میزان کارایی با استفاده از تکنیک ارزیابی ضربدری K بخشی، با K=۵ اندازه‌گیری شده است، به نحوی که مجموعه‌داده به پنج بخش به‌طور تقریبی مساوی تقسیم و هر یک از این زیرمجموعه‌ها یک بار به‌عنوان مجموعه آزمون[7] انتخاب می‌شود. به‌این ترتیب

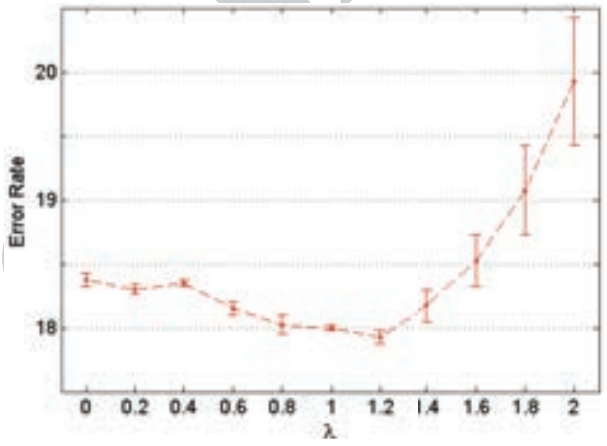
[6] Validation Set
[7] Test Set

Archive of SID
www.SID.ir

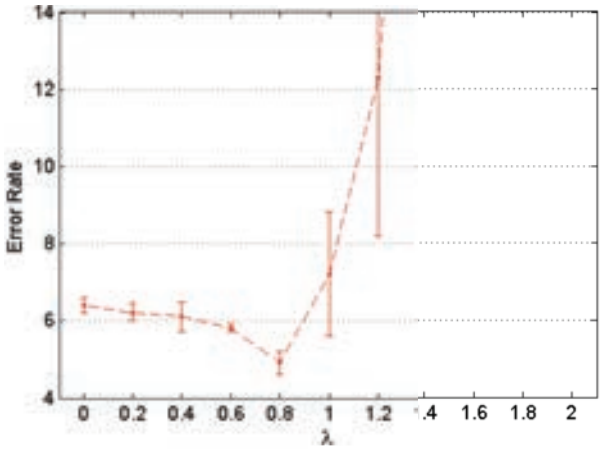
محاسبهٔ کارایی بر اساس میانگین این پنج بار اجرا، گزارش می‌شود. بر پایهٔ این تنظیمات و ملاحظات آزمایش‌های ذیل انجام شده است:

۴-۲-تأثیر پارامتر λ بر کارایی روش اختلاط خبره‌هایی با همبستگی منفی

در این آزمایش کارایی روش اختلاط خبره‌هایی با همبستگی منفی با ساختار پنج شبکهٔ پایه به‌ازای مقادیر مختلف پارامتر کنترلی λ در طبقه‌بندی مجموعه داده‌های وسائل نقلیه و تشخیص ارقام دست‌نوشته مورد آزمون قرار گرفته است. تأثیر تغییر پارامتر کنترلی λ در بازه [0:0.2:2] (از صفر تا دو با گام‌های دو دهم) بر روند تغییر نرخ خطای روش اختلاط خبره‌هایی با همبستگی منفی در (شکل ۳) نمایش داده شده است.



a(۳



b(۳

(شکل ۳): نمودارهای وابستگی نرخ خطای روش اختلاط خبره‌هایی با همبستگی منفی به پارامتر کنترلی λ به‌ترتیب در مسائل طبقه‌بندی مجموعه داده‌های تشخیص ارقام دست‌نوشته و وسائل نقلیه به ترتیب در نمودارهای b,a نشان داده شده است.

همان‌طور که در قبل اشاره شد، ضریب λ به روش اجازه می‌دهد تا با کنترل میزان تأثیر عبارت همبستگی، تعادلی شبه‌بهینه را در توازن عبارت‌های دقّت و گوناگونی ایجاد کند. هر دو نمودار نشان می‌دهد که در ابتدا نرخ خطای روش اختلاط خبره‌هایی با همبستگی منفی با افزایش مقدار λ کاهش پیدا کرده و سپس افزایش پیدا می‌کند. این چنین روند تغییر خطایی در این نمودارها نیز مسألهٔ توازن مذکور را نشان می‌دهد. همان‌طور که در نمودارها مشخّص است، روش ما در مورد مسأله‌های طبقه‌بندی وسائل نقلیه و تشخیص ارقام دست‌نویس به‌ترتیب در مقادیر λ=0.8 و λ=1.2 به حداقل نرخ خطای طبقه‌بندی رسیده است. هم‌چنین همان‌طور که در تصویر مشاهده می‌شود، با افزایش مقادیر λ، نرخ خطا به‌صورت چشم‌گیری افزایش پیدا می‌کند. این مسأله به این دلیل رخ داده است که در مقادیر بزرگ λ، روش به جای تمرکز بر کاهش نرخ خطای طبقه‌بندهای پایه، بیش‌تر عبارت همبستگی را کمینه می‌نماید.

۴-۳- روش اختلاط خبره‌هایی با همبستگی منفی در مقایسه با روش‌های پایه

در این آزمایش، روش اختلاط خبره‌هایی با همبستگی منفی با روش‌های پایهٔ سازندهٔ خود از جمله اختلاط خبره‌های پایه و یادگیری همبستگی منفی به‌ازای سیستم‌های ترکیبی با پنج شبکهٔ خبرهٔ پایه در پنج مسألهٔ آزمون طبقه‌بندی مورد بررسی و مقایسه قرار گرفته است. برای تنظیم بهینهٔ پارامتر λ در روش‌های یادگیری همبستگی منفی و اختلاط خبره‌هایی با همگرایی منفی، مقادیر مختلف λ در بازهٔ [0:0.1:2] روی مجموعه ارزیابی مورد آزمون قرار گرفته و مقدار بهینه انتخاب شده است. با این شرایط، نرخ خطای روش‌های ترکیبی مورد مقایسه در جدول ۲ گزارش شده است.

در این بررسی، روش آزمون آماری t-test بر اساس نرخ‌های خطای به‌دست آمده در جدول بالا مورد استفاده قرار گرفته تا تعیین کند آیا نرخ خطای به‌دست آمده توسط روش اختلاط خبره‌هایی با همگرایی منفی در مقایسه با روش‌های اختلاط خبره‌های پایه و یادگیری همبستگی منفی از لحاظ آماری به‌طور قابل توجّهی کاهش پیدا کرده است یا خیر. بر اساس این آزمون آماری با لحاظ‌کردن ٪۹۵ اطمینان‌پذیری، روش اختلاط خبره‌هایی با همگرایی منفی در تمام مسائل طبقه‌بندی مورد آزمایش، به‌جز مسألهٔ آواها، خطا را به‌طور قابل ملاحظه‌ای نسبت به روش اختلاط خبره‌های پایه افزایش داده است. در مسألهٔ طبقه‌بندی آواها،

Archive of SID
www.SID.ir

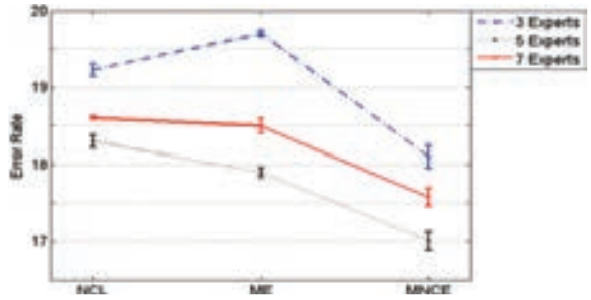
اگر چه روش ارائه شده در مقایسه با اختلاط خبره‌های پایه خطای کمتری داشته است، اما میزان این کاهش قابل ملاحظه نیست.

(جدول ۲): مقایسهٔ نرخ خطای روش ارائه شده، اختلاط خبره‌هایی با همبستگی منفی، در مقایسه با روش‌های پایهٔ سازندهٔ خود شامل اختلاط خبره‌های پایه و یادگیری همبستگی منفی.

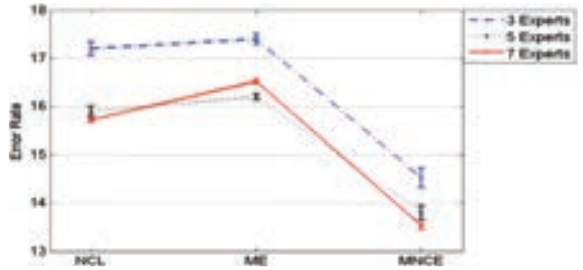
مجموعه-داده	نرخ خطا(٪)					
	اختلاط خبره‌های پایه		یادگیری همبستگی منفی		اختلاط خبره‌هایی با همبستگی منفی	
	میانگین (واریانس)	λ*	میانگین (واریانس)	λ*	میانگین (واریانس)	λ*
آواها	17.9 (9E-2)	0.8	18.3 (5E-2)	1	17.1 (2E-1)	
تصاویر ماهواره‌ای	16.2 (1E-1)	1	15.9 (7E-2)	1.1	13.8 (3E-1)	
ارقام دست-نویس	6.4 (2E-1)	1	5.5 (8E-2)	0.8	4.9 (3E-1)	
شیشه	30.8 (3E-1)	1	31.4 (1E-1)	0.1	28.3 (4E-1)	
وسائل نقلیه	19.5 (1E-1)	0.8	18.6 (6E-2)	1.2	17.7 (2E-1)	

بر اساس این آزمون آماری با لحاظ‌کردن ۹۵٪ اطمینان‌پذیری، روش اختلاط خبره‌هایی با همگرایی منفی در مقایسه با روش یادگیری همبستگی منفی کارایی را به‌طور قابل ملاحظه‌ای در همه مسائل مورد آزمون به‌جز مسائل آواها و تشخیص ارقام دست‌نوشته افزایش داده است. در این دو مسأله نیز اگرچه خطای طبقه‌بندی کاهش پیدا کرده، اما این کاهش از لحاظ این آزمون آماری قابل توجّه نبوده است.

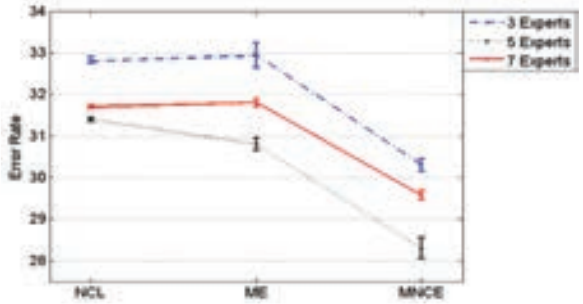
در قسمت دوم این آزمایش، برای بررسی نحوهٔ عملکرد این روش‌ها به‌ازای اندازه‌های مختلف سیستم‌های ترکیبی مورد مقایسه، به‌ازای سه حالت مختلف شامل ۳، ۵ و ۷ خبره پایه در پنج مسألهٔ آزمون طبقه‌بندی مورد آزمون قرار گرفته‌اند. با لحاظ‌کردن تأثیر اندازهٔ سیستم ترکیبی بر مبنای تعداد خبره‌های پایه در مقایسهٔ کارایی روش‌ها، می‌توانیم مقایسهٔ منطقی‌تری انجام دهیم. نرخ خطای روش اختلاط خبره‌هایی با همگرایی منفی در مقایسه با روش‌های اختلاط خبره‌هایی با همگرایی منفی و یادگیری همبستگی منفی در اندازه‌های مختلف سیستم ترکیبی در پنج مسألهٔ طبقه‌بندی در نمودارهای ذیل نمایش داده شده است:



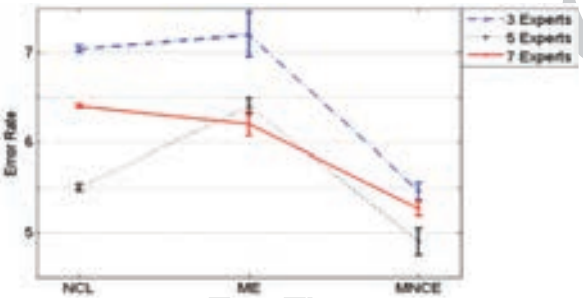
a (۴



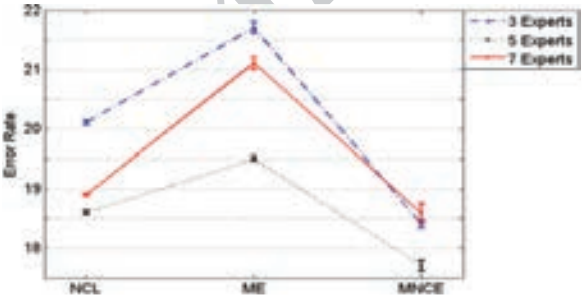
b (۴



c (۴



d (۴



e (۴

(شکل ۴): نمودارهای نرخ خطای اختلاط خبره‌هایی با همگرایی منفی در مقایسه با روش‌های اختلاط خبره‌های پایه و یادگیری همبستگی منفی، در اندازه‌های مختلف سیستم ترکیبی در پنج مسأله مشمول، به‌ترتیب مسأله‌های طبقه‌بندی آواها، تصاویر ماهواره‌ای، ارقام دست‌نویس، شیشه و وسائل نقلیه در تصاویر d-a نمایش داده شده است.

www.SID.ir

همان‌طور که در نمودارهای بالا مشاهده می‌شود، روش اختلاط خبره‌هایی با همگرایی منفی در مقایسه با روش‌های اختلاط خبره‌های پایه و یادگیری همبستگی منفی در تمام مسائل مورد آزمون و به‌ازای همهٔ اندازه‌های مختلف سیستم‌های ترکیبی نرخ خطای به مراتب کمتری داشته است.

۴-۴- بررسی اجزای خطای تعمیم در فرآیند یادگیری روش ارائه شده

جیکابز در مقاله *(Jacobs, 1997)* سیستم ترکیبی‌ای را بررسی کرده است که در آن خروجی کلی بر اساس میانگین وزن‌دار خبره‌های پایه محاسبه می‌شود و در چنین سیستمی، تخمین‌هایی برای اندازه‌گیری مقدار بایاس، واریانس و کوواریانس که سه جزء تشکیل دهندهٔ مربع خطای تعمیم هستند، ارائه شده است. برای ارائه این تخمین‌ها، ابتدا روابط پایه‌ای معرفی می‌شوند.

تخمین میانگین خروجی سیستم ترکیبی به‌ازای ورودی iاُم از مجموعه داده آزمون، طبق فرمول ذیل محاسبه می‌شود:

$$\overline{O_{ens}}(x_i)=\sum_{n=1}^{N}O_{ens}(x_i,n) \qquad (۱۶)$$

که $O_{ens}(x_i,n)$ خروجی سیستم ترکیبی به‌ازای ورودی x_i در اجرای nاُم و N تعداد کل مراحل اجراست.

تخمین میانگین وزن‌دار خروجی خبره kاُم به‌ازای ورودی x_i طبق فرمول ذیل محاسبه می‌شود:

$$\overline{g_k(x_i)O_k(x_i)}=\frac{1}{N}\sum_{n=1}^{N}g_k(x_i,n)O_k(x_i,n) \qquad (۱۷)$$

که $g_k(x_i,n)$ خروجی kاُم شبکهٔ میانجی در اجرای nاُم و $O_k(x_i,n)$ خروجی خبره kاُم در اجرای nاُم است.

بر این اساس، مقادیر تخمینی بایاس، واریانس و کوواریانس طبق روابط ذیل محاسبه می‌شوند:

$$IB=\frac{1}{|X|}\sum_{i=1}^{|X|}\left(\overline{O_{ens}}(x_i)-y_i\right)^2 \qquad (۱۸)$$

$$IV=\sum_{k}\frac{1}{|X|}\sum_{i=1}^{|X|}\frac{1}{N}\sum_{n=1}^{N}\left(g_k(x_i,n)O_k(x_i,n)-\overline{g_k(x_i)O_k(x_i)}\right)^2 \qquad (۱۹)$$

$$IC=\sum_{k}\sum_{m\neq k}\frac{1}{|X|}\sum_{i=1}^{|X|}\frac{1}{N}\sum_{n=1}^{N}\frac{\left(g_k(x_i,n)O_k(x_i,n)-\overline{g_k(x_i)O_k(x_i)}\right)\cdot}{\left(g_m(x_i,n)O_m(x_i,n)-\overline{g_m(x_i)O_m(x_i)}\right)} \qquad (۲۰)$$

که y_i برچسب متناظر ورودی iاُم و $|X|$ اندازهٔ مجموعهٔ آزمون است.

مقدار تخمینی میانگین مربع خطا نیز که جمع سه عبارت بالایی است، طبق رابطهٔ ذیل محاسبه می‌شود:

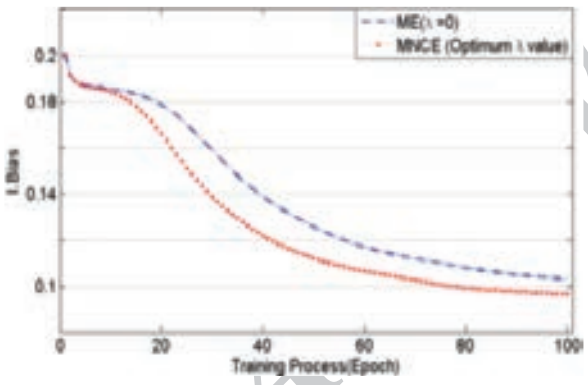
$$IMSE=\frac{1}{|X|}\sum_{i=1}^{|X|}\frac{1}{N}\sum_{n=1}^{N}\left(O_{ens}(x_i,n)-y_i\right)^2 \qquad (۲۱)$$

ما روند تغییر عبارت‌های مختلف تشکیل‌دهندهٔ خطای تعمیم در فرآیند یادگیری روش‌های اختلاط خبره‌های با همگرایی منفی و اختلاط خبره‌های پایه را در یک مسأله رگرسیون بررسی کرده‌ایم. تابع مورد آزمون رگرسیون در این مسأله عبارتند از:

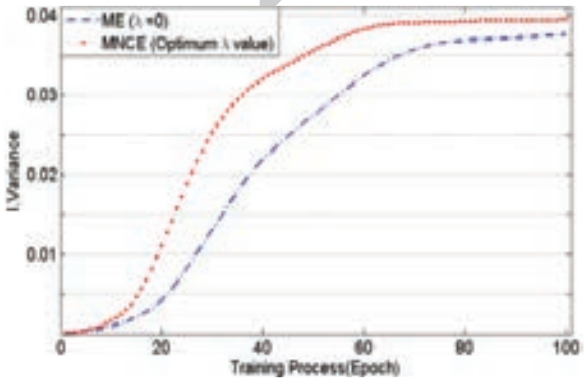
$$f(x)=\frac{1}{13}\left[10\sin(\pi x_1x_2)+20\left(x_3-\frac{1}{2}\right)^2+10x_4+5x_5\right]-1 \qquad (۲۲)$$

که $x=[x_1\ x_2\ x_3\ x_4\ x_5]$ که مقادیر هر بعد در بازه [0,1] و مقدار خروجی تابع نیز در بازهٔ [0,1] تغییر می‌کند.

شرایط و تنظیمات آزمایش مشابه پیشنهاد جیکابز *(Jacobs, 1997)* لحاظ شده است. روند تغییرات عبارت‌های تخمینی بایاس، واریانس، کوواریانس و میانگین مربع خطا در طی روند یادگیری روش‌های اختلاط خبره‌های با همگرایی منفی و اختلاط خبره‌های پایه در نمودارهای ۵-a) (d به‌ترتیب نمایش داده شده است.

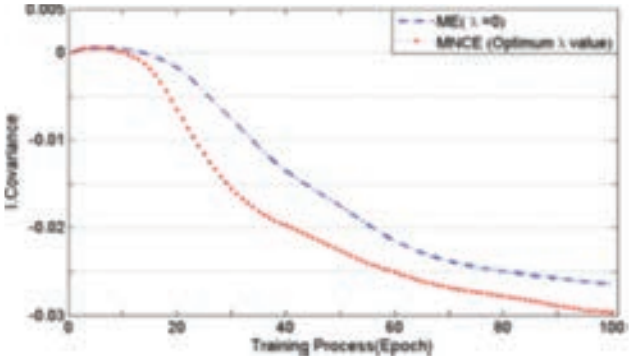


a (۵

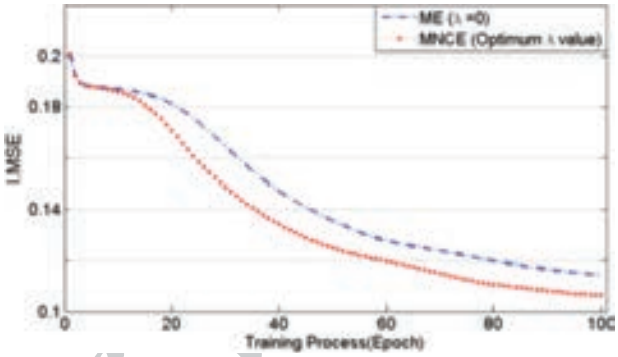


b (۵

Archive of SID
www.SID.ir



c (۵



d (۵

نمودار ۵ (a-d) روند تغییر عبارت‌های بایاس، واریانس، کوواریانس و میانگین مربع خطا در طی روند یادگیری روش‌های اختلاط خبره‌های با همگرایی منفی و اختلاط خبره‌های پایه که به‌ترتیب با رنگ‌های آبی و قرمز نشان داده شده است.

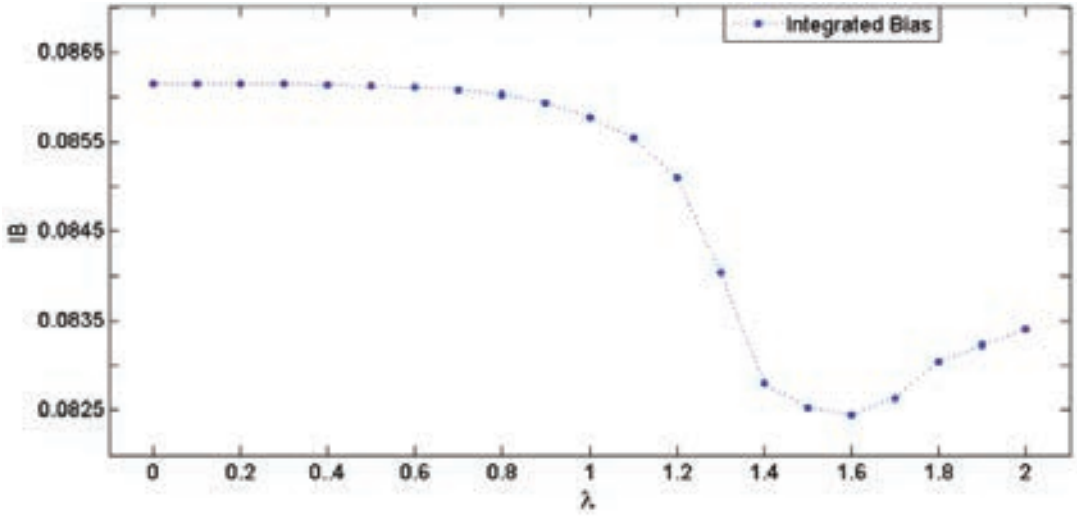
همان‌طور که نتایج آزمایش در نمودارها نشان می‌دهد، مقدار عمدهٔ میانگین مربع خطا از بایاس ناشی می‌شود. به همین دلیل، همان‌طور که در نمودارها مشاهده می‌شود، روند تغییر میانگین مربع خطا بیش از همه، به روند تغییر عبارت بایاس شباهت دارد.

خبره‌های پایه در روش‌های اختلاط خبره‌های با همگرایی منفی و اختلاط خبره‌های پایه، طی فرآیند یادگیری در توزیع‌های مختلفی از فضای مسأله به‌صورت محلّی خبره می‌شوند. همان‌طور که جیکابز *(Jacobs, 1997)* اشاره کرده و نیز در تصاویر b و c مشخّص است، در طی فرآیند محلّی‌سازی خبره‌ها مقدار واریانس تخمینی خروجی‌های وزن‌دار خبره‌ها افزایش و مقدار کوواریانس تخمینی خروجی‌های وزن‌دار خبره‌ها کاهش پیدا کرده است.

همان‌طور که در نمودار بالا مشخص است، پارامتر منظم‌سازی اضافه شده به تابع خطای روش ارائه شده، علاوه‌بر کنترل عبارت‌های واریانس و کوواریانس، مقدار عبارت بایاس را نیز در مقایسه با روش اختلاط خبره‌ها کاهش داده است. همان‌طور که در نمودار c مشخّص است، بر اثر عبارت اضافه شدهٔ کنترل همبستگی منفی در تابع خطای روش اختلاط خبره‌های با همگرایی منفی، عبارت کوواریانسِ روش ما در مقایسه با روش اختلاط خبره‌ها کاهش پیدا کرده است؛ بنابراین، می‌توان گفت که روش اختلاط خبره‌های با همگرایی منفی با اعمال همبستگی منفی بیشتر میان خبره‌های پایه، به‌رغم افزایش واریانس موجب کاهش دو عبارت بایاس و کوواریانس خطا شده که درمجموع کاهش میانگین مربع خطای تعمیم را نتیجه داده است.

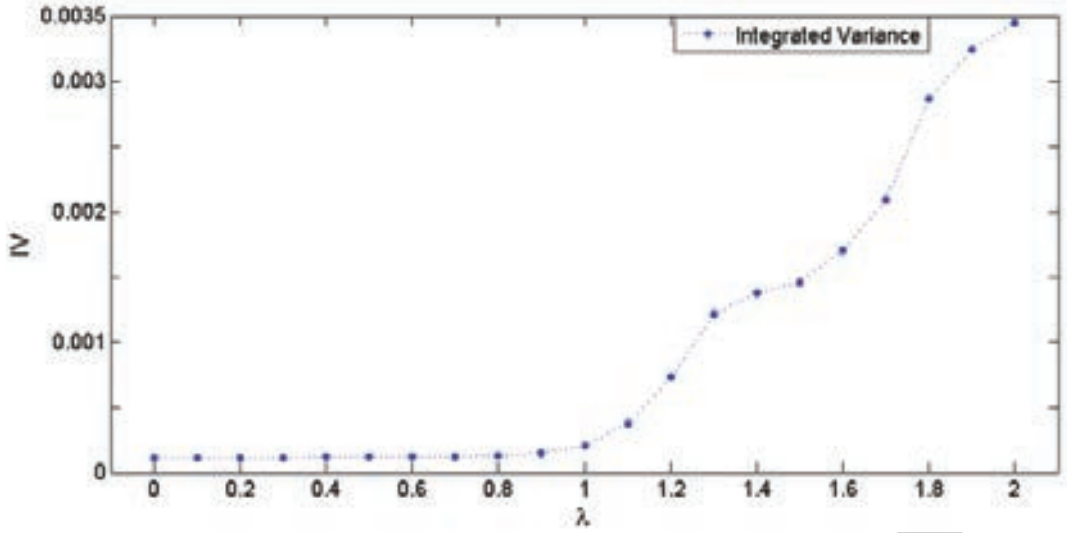
۴-۴-۱- تأثیر پارامتر λ بر عبارت‌های خطای تعمیم در روش ارائه شده

هدف از این آزمایش بررسی وابستگی عبارت‌های بایانس، واریانس، کوواریانس و خطای تعمیم به پارامتر کنترلی λ در روش ارائه شده است. شرایط و تنظیمات آزمایش، مشابه آزمایش قبلی لحاظ شده است، با این تفاوت که مقادیر مورد بررسی، پس از اتمام فرآیند یادگیری ترکیب اندازه‌گیری شده است. نتایج حاصل از روش اختلاط خبره‌های با همگرایی منفی در حل مسألهٔ رگرسیون مذکور در آزمایش قبل به‌ازای مقادیر مختلف λ در بازه [0:0.1:2] در نمودارهای ذیل نمایش داده شده است.

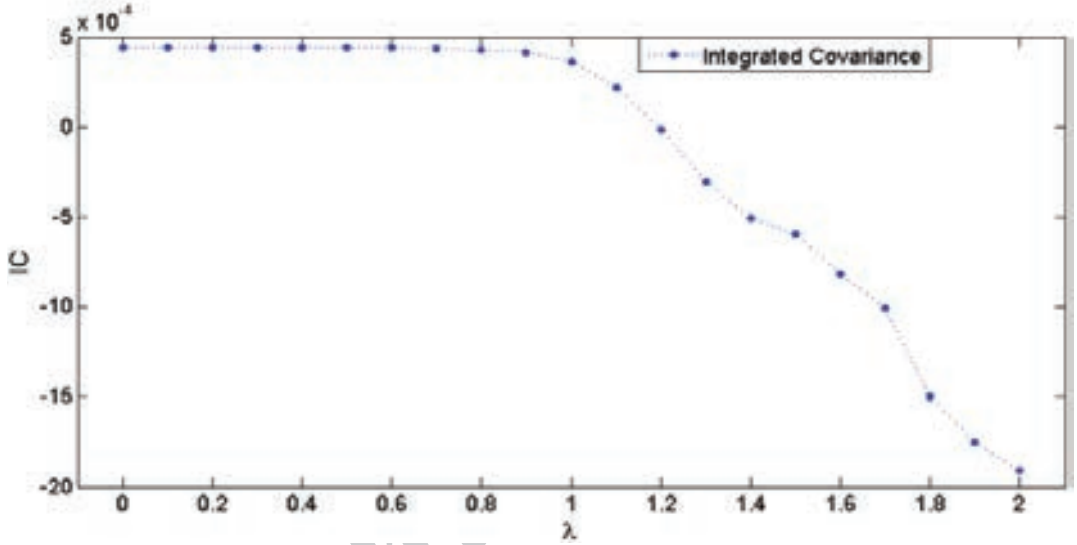


a (۶

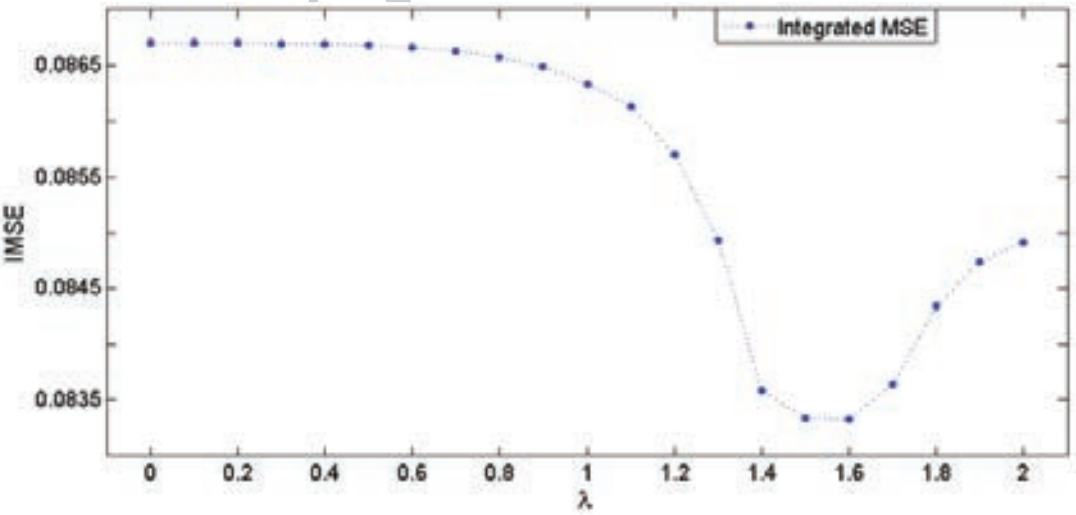
Archive of SID
www.SID.ir



b (۶



c (۶



d (۶

(شکل۶ a-d): وابستگی عبارت‌های بایانس، واریانس، کوواریانس و خطای تعمیم به پارامتر کنترلی λ در روش اختلاط خبره‌های با همگرایی منفی.

www.SPIJ.ir

نتایج نشان می دهد که به‌ازای تغییر λ در بازهٔ [0,2]، ابتدا عبارت بایاس کاهش و سپس افزایش پیدا کرده است، به‌نحوی که کمینهٔ مقدار بایاس در λ=1.6 حاصل شده است. همان‌طور که نمودار a نشان می‌دهد، به‌ازای مقادیر بزرگ λ عبارت بایاس افزایش پیدا می‌کند. این پدیده به این دلیل است که در چنین شرایطی فرآیند یادگیری به‌جای کمینه‌کردن عبارت خطای ترکیب، بر روی کمینه‌کردن عبارت همبستگی خبره‌های پایه تمرکز می‌کند.

همان‌طور که ذکر شد، با افزایش پارامتر λ فرآیند یادگیری بر روی کاهش عبارت خطای همبستگی تمرکز می‌کند که به همین دلیل با افزایش λ مقدار واریانس افزایش و مقدار کوواریانس کاهش پیدا کرده است. کاهش عبارت کوواریانس به حدّی است که منجربه به وجود آمدن کوواریانس منفی میان خبره‌های پایه شده است.

مشابه آزمایش قبلی، نتایج نشان می‌دهد که عمده خطای تعمیم از عبارت بایاس ناشی می‌شود؛ به‌نحوی که روند تغییر خطای تعمیم، بیش از همه به روند تغییر بایاس شباهت دارد. نکتهٔ قابل توجّه این است که روش ارائه شده، نه تنها میزان واریانس و کوواریانس سامانهٔ ترکیبی را کنترل می‌کند، بلکه مقدار بایاس آن را نیز به‌طور قابل توجّهی کاهش داده است. بنابراین روش ارائه شده بر مبنای کنترل تمام عبارت‌های خطا به‌وسیلهٔ انتخاب مقدار بهینهٔ پارامتر λ، موجب افزایش قابلیت تعمیم شده است. قابل ذکر است که در این سیستم ترکیبی ارائه شده، توازن بایاس-واریانس-کوواریانس به‌ازای مقدار λ=۱/۶ به مقدار بهینهٔ خود رسیده که کمینه‌شدن خطای تعمیم را نتیجه داده است.

۵- نتیجه‌گیری

در این مقاله، ویژگی‌های مختلف دو روش اختلاط خبره‌ها و یادگیری همبستگی منفی را مرور و نقاط قوّت و ضعف آن‌ها را در مقایسه با یکدیگر بررسی کردیم. بررسی خصوصیات این دو روش در مقایسه با یکدیگر نشان داد که آنها ویژگی‌های متفاوت و البّته مکملی نسبت به هم دارند؛ به‌نحوی که اگر بتوان سیستمی ترکیبی شامل ویژگی‌های مثبت هر دو روش طراحی کرد، به‌احتمال کارایی بهتر از روش‌های پایهٔ خود داشته باشد. در این مقاله، ایده‌ای را برای ترکیب ویژگی‌های این دو روش پیشنهاد کرده‌ایم: ایدهٔ اختلاط خبره‌هایی با همگرایی منفی. در روش ارائه شده، قابلیت پارامتر کنترلی روش یادگیری همبستگی منفی به تابع خطای روش اختلاط خبره‌ها افزوده شده، که روش اختلاط خبره‌ها را قادر می‌سازد تعادل بهینه‌ای را در توازن بایاس- واریانس-کوواریانس ایجاد و کارایی را افزایش دهد. روش ترکیبی اختلاط خبره‌هایی با همگرایی منفی، در چند مسألهٔ آزمون پیش‌بینی و طبقه‌بندی با روش‌های پایهٔ اختلاط خبره‌ها و یادگیری همبستگی منفی مورد مقایسه قرار گرفته است. نتایج آزمایش‌ها نشان داده است که روش پیشنهادی، با حفظ نقاط قوّت و کاهش ضعف‌های روش‌های پایه، توانسته کارایی را به‌طور قابل ملاحظه‌ای نسبت به آنها افزایش دهند. همچنین در آزمایش پایانی با بررسی پارامترهای میانگین مربع خطا، شامل: بایاس، واریانس و کوواریانس، علّت برتری اختلاط خبره‌هایی با همگرایی منفی بر اختلاط خبره‌های پایه را توجیه و نمایش داده‌ایم.

سپاس گزاری

این تحقیق در قالب طرح پژوهشی "افزایش نرخ کارایی طبقه‌بندی با استفاده از تجمیع ویژگی‌های مؤثّر روش‌های مختلف ترکیب شبکه‌های عصبی" با حمایت دانشگاه آزاد اسلامی واحد کرج انجام پذیرفته شده است.

۶- منابع

Avnimelech R, Intrator N (1999) Boosted mixture of experts: an ensemble learning scheme. Neural computation 11 (2):483-497.

Brown G, Wyatt JM (2003) Negative correlation learning and the ambiguity family of ensemble methods. Multiple Classifier Systems, Proceeding 2709:266-275.

Ebrahimpour R, Nikoo H, Masoudnia S, Yousefi MR, Ghaemi MS (2011) Mixture of MLP-experts for trend forecasting of time series: A case study of the Tehran stock exchange. International Journal of Forecasting 27 (3):804-816. doi:10.1016/j.ijforecast.2010.02.015.

Ghaemi M, Masoudnia S, Ebrahimpour R (2010) A new framework for small sample size face recognition based on weighted multiple decision templates. Neural Information Processing Theory and Algorithms 6643/2010:470-477. doi:10.1007/978-3-642-17537-4_58.

Hansen JV (2000) Combining predictors: Meta machine learning methods and bias/variance & ambiguity decompositions. Computer Science Dept., Aarhus Univ.

Islam MM, Yao X, Nirjon SMS, Islam MA, Murase K (2008a) Bagging and boosting negatively correlated neural networks. Ieee T Syst Man Cy B 38 (3):771-784. doi:Doi 10.1109/Tsmcb.2008.922055.

Archive of SID

www.SID.ir

Islam MM, Yao X, Shahriar Nirjon S, Islam MA, Murase K (2008b) Bagging and boosting negatively correlated neural networks. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 38 (3):771-784.

Jacobs RA (1997) Bias/variance analyses of mixtures-of-experts architectures. Neural Comput 9 (2):369-383.

Jacobs RA, Jordan MI, Nowlan SJ, Hinton GE (1991) Adaptive mixtures of local experts. Neural computation 3 (1):79-87.

Kuncheva LI (2004) Combining pattern classifiers: methods and algorithms. Wiley-Interscience.

Liu Y, Yao X (1999a) Ensemble learning via negative correlation. Neural Networks 12 (10):1399-1404.

Liu Y, Yao X (1999b) Simultaneous training of negatively correlated neural networks in an ensemble. Ieee T Syst Man Cy B 29 (6):716-725.

Masoudnia S, Ebrahimpour R (2012) Mixture of Experts, a Literature Survey. Artif Intell Rev. doi:Accepted.

Masoudnia S, Rostami M, Tabassian M, Sajedin A, Ebrahimpour R Evidence-based mixture of MLP-experts. In: IEEE International Joint Conference of Neural Networks, 2010 2010. IEEE, pp 1-7.

Rokach L (2010a) Ensemble-based classifiers. Artificial Intelligence Review 33 (1):1-39.

Rokach L (2010b) Pattern classification using ensemble methods, vol 75. World Scientific Pub Co Inc,.

Ubeyli ED (2009) Modified mixture of experts employing eigenvector methods and Lyapunov exponents for analysis of electroencephalogram signals. Expert Syst 26 (4):339-354. doi:DOI 10.1111/j.1468-0394.2009.00490.x.

Ubeyli ED, Ilbay K, Ilbay G, Sahin D, Akansel G (2010) Differentiation of Two Subtypes of Adult Hydrocephalus by Mixture of Experts. J Med Syst 34 (3):281-290. doi:DOI 10.1007/s10916-008-9239-4.

Waterhouse SR (1997) Classification and regression using mixtures of experts. Unpublished doctoral dissertation, Cambridge University

جـــواد محمّـــدزاده کارشناســی و کارشناسی ارشد خـود را در رشـتهٔ علـوم کامپیوتر به‌ترتیب از دانشگاه شهید بـاهنر کرمان در سال ۱۳۸۳ و دانشگاه تهران در سـال ۱۳۸۶ دریافـت کـرده اسـت. هـم اکنون ایشان عضو هیأت علمـی گـروه کـامپیوتر در دانشـگاه آزاد واحد کـرج هسـتند و در زمینـه‌هـای علمـی طراحـی و بهینه‌سازی الگوریتم‌ها، محاسبات نـرم، شـبکه‌هـای عصـبی، بیوانفورماتیـک، طراحـی سـاختار دوم RNA و مـدل‌سـازی شبکه‌های تنظیم ژنی فعّالیت می‌کنند.

نشانی رایانامک ایشان، عبارتست از:

j.mohammadzadeh@kiau.ac.ir

ســـعید مســـعودنیا تحصـــیلات دورهٔ کارشناسی و کارشناسی ارشـد را در رشـتهٔ علوم کامپیوتر در دانشگاه تهران بـه‌ترتیـب در سـال‌هـای ۱۳۸۷ و ۱۳۹۰ بـه پایـان رسانده است. زمینه‌هـای تحقیقـاتی مـورد علاقهٔ ایشان عبارتنـد از: یـادگیری ماشـین، پـردازش الگـو و بینایی ماشین و آنالیز سیگنال‌های مغزی است.

نشانی رایانامک ایشان، عبارتست از:

s.masoudnia@gmail.com

علی آرانی تحصیلات دورهٔ کارشناسی و کارشناسی ارشـد را در رشـتهٔ مهندسـی برق در دانشگاه تربیت دبیر شهید رجایی تهران بـه‌ترتیـب در سـال‌هـای ۱۳۸۸ و ۱۳۹۰ به پایان رسانده است. زمینه‌هـای تحقیقـاتی مـورد علاقـهٔ ایشـان عبارتنـد از: پـردازش الگـو، شبکه‌های عصبی و علوم اعصاب شناختی است.

نشانی رایانامک ایشان، عبارتست از:

ali.araany@gmail.com

Archive of SID