

روش پیش‌تعلیم لایه‌به‌لایه دوسویه برای تعلیم شبکه‌های عصبی عمیق

سیده زهره سید صالحی^۱ و سید علی سید صالحی^۲

۱- دانشجوی دکتری، دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر، تهران، ایران

z.seyyedsalehi@aut.ac.ir

۲- دانشیار، دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر، تهران، ایران

ssalehi@aut.ac.ir

چکیده: در این مقاله، یک روش پیش‌تعلیم دوسویه برای همگرا نمودن تعلیم شبکه‌های عصبی عمیق با یادگیری دیگرانجمنی ارائه شده است. تعلیم این شبکه‌ها به دلیل مواجهه بودن با تعداد بالای کمینه‌های موضعی اغلب همگرا نمی‌گردد. این درحالیست که با مقداردهی اولیه مناسب وزن‌های شبکه، می‌توان از بسیاری از کمینه‌های موضعی اجتناب نمود. روش پیش‌تعلیم لایه‌به‌لایه دوسویه روشی سریع و کارا می‌باشد که در یک مسیر دوسویه به‌طور جلوسو و عقب‌سو با استفاده از ورودی‌ها و خروجی‌های مطلوب شبکه، به تنظیم مقادیر اولیه وزن‌های آن می‌پردازد. برای این منظور از تعلیم شبکه‌های کمکی یک لایه پنهان مبتنی بر وزن‌های لایه تحت پیش‌تعلیم از شبکه عمیق و وزن‌های کمکی استفاده می‌شود. سپس مقادیر وزن حاصل از تعلیم اینها در ساختار اصلی شبکه تحت پیش‌تعلیم قرار داده می‌شوند و برای تنظیم دقیق وزن‌ها، تعلیم یکپارچه صورت می‌گیرد. این روش برای پیش‌تعلیم وزن‌های سه شبکه عصبی عمیق بازناسن فرد، حالت‌های احساسی و ارقام دست‌نوشته مورد استفاده قرار گرفت و نشان داده شد که با به‌کارگیری این روش پیش‌تعلیم، سرعت همگرایی تعلیم به‌طور چشمگیری افزایش می‌یابد. همچنین میزان بازناسنی‌ها در پایگاه داده‌های چهره به میزان قابل توجهی بهبود می‌یابد که حاکی از افزایش قدرت تعمیم شبکه با استفاده از این روش می‌باشد.

واژه‌های کلیدی: پیش‌تعلیم، دوسویه، دیگرانجمنی، ساختار عمیق، شبکه‌های عصبی چند لایه، همگرایی تعلیم

۱- مقدمه

سلسله‌مراتبی مؤلفه‌ها در آن‌ها فراهم می‌شود. به‌گونه‌ای که مؤلفه‌های سطح بالا به‌صورت ترکیبی از مؤلفه‌های سطح پایین در چندین سطح تشکیل می‌گردند [۳]. این درحالیست که تعلیم معمول شبکه‌های با بیش از دو لایه پنهان، اغلب به نتایج ضعیفی منجر می‌شود. به‌طوریکه وقتی تلاش می‌شود تمام لایه‌ها از طریق یک تابع معیار مانند میزان شباهت به ورودی‌ها یا شباهت به طبقه‌های مطلوب، تعلیم داده شوند نتایج ممکن است بدتر از مدل‌های کم عمق باشد [۴]. این به دلیل افزایش تعداد کمینه‌های موضعی فعال در این ساختارها می‌باشد که به نوع توابع و مقداردهی اولیه پارامترهای شبکه بستگی زیاد دارند [۵]. به‌عبارت دیگر در تعلیم ساختارهای عمیق شبکه، با افزایش تعداد لایه‌ها و نوروها، مشکل کمینه‌های موضعی محسوس‌تر می‌شود

در چند سال اخیر با توجه به شواهد موجود مبنی بر حضور ساختارهای عمیق در مغز انسان [۱]، محققین به سمت تعلیم شبکه‌های عصبی چند لایه عمیق گرایش پیدا کرده‌اند. این ساختارها قادر به یادگیری چندین سطح بازنمایی از ورودی می‌باشند [۲]. لذا امکان استخراج

^۱ تاریخ ارسال مقاله : ۱۳۹۲/۰۱/۱۸

تاریخ پذیرش مقاله : ۱۳۹۳/۱۲/۰۶

نام نویسنده مسئول : سیده زهره سید صالحی

نشانی نویسنده مسئول : ایران - تهران - دانشگاه صنعتی امیرکبیر

- دانشکده دانشکده مهندسی پزشکی

وزن‌های شبکه عصبی عمیق‌تر نسبت به مقادیر مطلوب نهایی آن‌ها در وضعیتی قرار می‌گیرند که به آن‌ها نزدیک‌تر و به عبارتی مشرف به هدف نهایی باشند. نشان داده شده است که این روش قادر به برطرف کردن کمینه‌های موضعی در مسیر شبکه‌های عصبی جلوسوی عمیق می‌باشد. در سال ۲۰۰۶، هیتون^۲ روش تجزیه به ماشین‌های بولتزمان محدود شده^۳ (RBM) را برای پیش‌تعلیم شبکه‌های عصبی چند لایه به‌منظور کاهش بعد غیرخطی ارائه داد. در این روش شبکه چندلایه به تعداد متناظری RBM شکسته می‌شود و پیش‌تعلیم وزن‌ها از طریق این RBM‌ها صورت می‌گیرد [۱۵]. در [۹۸] نیز روش پیش‌تعلیم لایه‌به‌لایه برای پیش‌تعلیم شبکه‌های خودانجمنی گلوگه عمیق به‌منظور استخراج مولفه‌های اساسی آورده شده است که در این مقاله نسخه دوسویه آن برای پیش‌تعلیم شبکه‌های عصبی عمیق با کاربردهای دیگرانجمنی ارائه شده است. با توجه به اینکه روش‌های پیش‌تعلیم موجود برای یادگیری‌های خودانجمنی ارائه شده‌اند، این روش به‌عنوان راه‌حلی نو برای یادگیری دیگرانجمنی شبکه‌های عصبی عمیق می‌باشد. روش پیش‌تعلیم لایه‌به‌لایه دوسویه که حالت کلی‌تری از روش پیش‌تعلیم لایه‌به‌لایه [۹] است، به‌عنوان یک روش پیش‌تعلیم باسرپرستی با انتخاب مناسب شرایط اولیه، الگوریتم پس‌انتشار خطا را در جهت رسیدن به کمینه مطلوب کمک می‌کند. روش ارائه شده برای پیش‌تعلیم شبکه‌های عصبی عمیق با هدف بازشناسی تصاویر مورد ارزیابی قرار گرفته است.

در ادامه به معرفی روش پیش‌تعلیم لایه‌به‌لایه دوسویه پرداخته می‌شود. سپس در بخش سوم پیاده‌سازی‌ها و نتایج آن‌ها آورده شده است. در نهایت جمع‌بندی و نتیجه‌گیری در بخش چهارم ارائه می‌شود.

۲- روش پیش‌تعلیم لایه‌به‌لایه دوسویه

شبکه‌های عصبی با ساختار عمیق با استفاده از لایه‌های متعددی از نورون‌ها، ظرفیت یادگیری تبدیلات غیرخطی چندمرحله‌ای و پیچیده را دارا هستند. اما همان‌طور که عنوان

[۶]. لذا مقداردهی اولیه مناسب وزن‌های شبکه جهت همگرا نمودن تعلیم آن ضروری به‌نظر می‌رسد، به‌طوریکه ممکن است به چندین مرحله متوالی از پیش‌تعلیم نیاز شود به‌گونه‌ای که هر مرحله پیش‌تعلیم، به‌عنوان پیش‌نیاز برای مرحله تعلیم بعدی، به‌منظور عبور دادن فرایند تعلیم از موانع کمینه‌های موضعی موجود در میان راه، مورد نیاز باشد. بنابراین در کنار رویکرد مستقیم و معمول تعلیم یک شبکه عصبی چند لایه که از روش پس‌انتشار خطا استفاده می‌کند، رویکردهای جدیدی وجود دارند که با بالا رفتن پیچیدگی داده و تعداد نورون‌ها، مفید واقع می‌شوند که تحت عنوان روش‌های پیش‌تعلیم مطرح هستند [۷۸].

اساساً روش‌های پیش‌تعلیم با هدف رهاسازی فرآیند تعلیم از کمینه‌های موضعی موجود در میان راه به‌عنوان یک مانع اساسی در روند آموزش، مورد استفاده قرار می‌گیرند. این روش‌ها به‌دنبال یافتن نقطه شروع مناسبی برای وزن‌های شبکه می‌باشند. روش‌های پیش‌تعلیم علاوه بر اینکه فرایند همگرایی تعلیم را تسهیل و تسریع می‌کنند، قدرت تعمیم شبکه را نیز بهبود می‌دهند. در [۲۹] نشان داده شده است که با خطای تعلیم مساوی، شبکه‌ای که قبل از تعلیم، مقداردهی اولیه مناسب شده است، خطای آزمون پایین‌تری دارد. به‌عبارتی پیش‌تعلیم، پارامترهای شبکه را وادار می‌کند تا در منطقه‌ای از فضای پارامترها که جواب مطلوب در آن است، قرار بگیرند [۲۰، ۱۱].

یکی از روش‌های موجود جهت پیش‌تعلیم شبکه‌های عمیق روش مشرف به هدف می‌باشد. این روش اولین بار جهت تعلیم شبکه‌های عصبی عمیق در [۱۲] برای استخراج مؤلفه‌های اساسی غیرخطی استفاده شد و سپس در [۱۴]، [۱۳] نیز به‌کار رفت. در این روش ابتدا شبکه عصبی مورد استفاده یک شبکه عصبی گلوگه^۱ (BNN) یک لایه پنهان فرض می‌شود که در طی آموزش، به عمق آن افزوده می‌شود. در هر مرحله از رشد تعداد لایه‌ها و عمق شبکه، مقادیر تعلیم‌یافته وزن‌های شبکه کم‌عمق‌تر قبلی، به‌عنوان مقادیر اولیه پیش‌تعلیم‌یافته برای شبکه عمیق‌تر مرحله بعد مورد استفاده قرار می‌گیرند. بدین طریق مقادیر اولیه

مختلف، در محل مطلوب از نظر تمایزدهی نمایش الگوها در لایه‌های متوالی می‌باشند. در نتیجه این مطلب است که در خروجی نیز تمایزدهی الگوها تا حد زیادی کاهش می‌یابد. در روش پیش‌تعلیم لایه‌به‌لایه دوسویه با قرار دادن مرحله‌به‌مرحله ابرصفحات نورون‌ها در محل مناسب برای حفظ تمایزدهی در لایه‌ها، از ایجاد مشکل فوق‌جلوگیری می‌شود. لذا می‌توان با شکستن مسئله پیچیده یادگیری شبکه‌های عصبی عمیق به چند زیرمسئله یادگیری ساختارهای ساده‌تر و حل مسئله همگرایی آن‌ها و سپس استفاده از آن‌ها به‌عنوان مقادیر اولیه برای شبکه عصبی عمیق، تعلیم خروجی مطلوب را برای شبکه ساده نمود. در این روش مستقیماً از اطلاعات ورودی و خروجی مطلوب در تنظیم اولیه وزن‌های شبکه عمیق استفاده می‌شود. مقادیر اولیه وزن‌های لایه‌های مختلف مبتنی بر ورودی یا خروجی برحسب اینکه آن لایه به کدام نزدیکتر باشد، تعیین می‌شوند. در این مراحل با توجه به اینکه شبکه‌های یک لایه پنهان تعلیم داده می‌شوند، تعلیم سریع و نقش کمینه‌های موضعی کمرنگ‌تر است. این روش پیش‌تعلیم، وزن‌های اولیه شبکه عصبی عمیق را به موقعیتی منتقل می‌کند که از بسیاری از کمینه‌های موضعی ممکن رهیده باشد.

از این طریق وزن‌های نورون‌های لایه‌های مختلف شبکه از یک روش پیش‌تعلیم نسبتاً کارآیی بهره‌مند می‌شوند. باید در نظر داشت که این روش پیش‌تعلیم لایه‌به‌لایه برای هر شبکه جلوسو با هر تعداد لایه پنهان قابل استفاده است. می‌توان نشان داد که روش ارائه شده برای پیش‌تعلیم وزن‌ها روش بسیار کارآیی است.

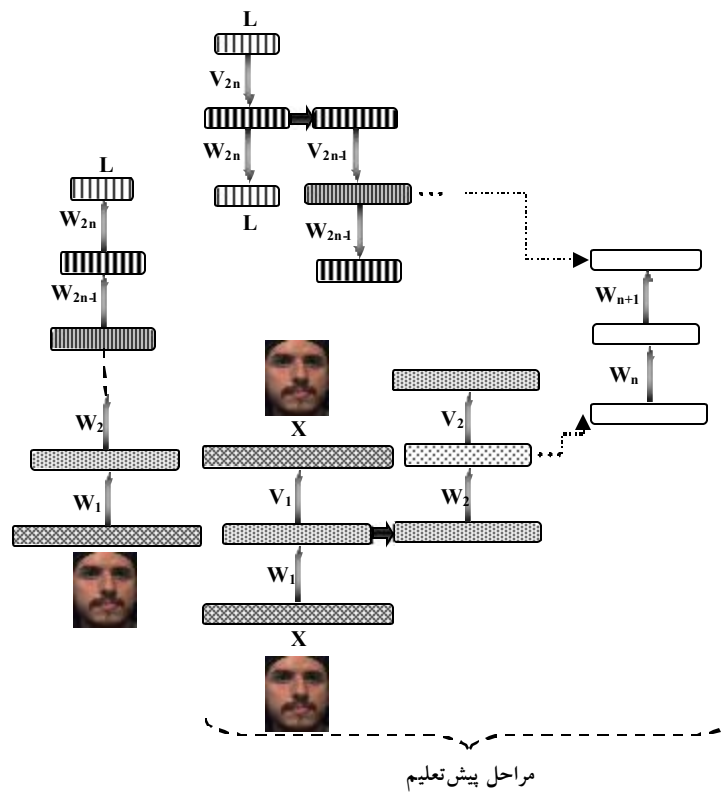
در روش پیش‌تعلیم لایه‌به‌لایه دوسویه، اگر شبکه عمیق مطابق شکل (۱) داری $2n$ لایه وزن باشد، تجزیه ساختار آن به‌طور موازی از ابتدا و انتهای به‌ترتیب به‌طور جلوسو و عقب‌سو صورت می‌گیرد. به این‌صورت که در مسیر جلوسو BNN‌های یک لایه پنهانی تعریف می‌شوند که وزن‌های ورودی آن‌ها (W_i) بیانگر وزن‌های لایه موردنظر از شبکه عمیق باشند. وزن‌های خروجی آن‌ها (V_i) نیز وزن‌های کمکی هستند که برای تعیین W_i مورد استفاده قرار می‌گیرند. در این مسیر، BNN اول با بردار ورودی و BNN‌های بعدی با تصویر بردار ورودی BNN قبلی در

شد با افزایش تعداد لایه‌های شبکه‌های عصبی، همگرا نمودن تعلیم آن‌ها توسط الگوریتم پس‌انتشار خطا به‌دلیل وجود کمینه‌های موضعی، مشکل و گاهاً غیرممکن می‌شود. حال اگر بتوان به‌طریقی تعداد کمینه‌های موضعی ممکن را در مسیر تعلیم شبکه محدود نمود، همگرایی بهبود می‌یابد. باید توجه داشت که شبکه‌های عصبی عمیق در صورتی قادر به تولید خروجی‌ها با دقت مطلوبی می‌باشند که بتوانند در تجزیه و تحلیل اطلاعات ورودی در لایه‌های مختلف، اتلاف اطلاعات نداشته باشند. لذا لازم است ورودی‌های متفاوت را در لایه‌های مختلف خود، به‌صورت متمایزی بیان نمایند. در صورتیکه دو نمونه متمایز از داده‌های ورودی، در لایه‌های شبکه با یک بیان واحد و غیرمتمایز از یکدیگر بازنمایی شوند، آنگاه تبدیل غیرخطی ورودی تا این لایه، نتوانسته برای این نمونه‌ها به‌صورت یک‌به‌یک و متمایزگرانه عمل نماید. لذا بخشی از اطلاعات ورودی از دست رفته است.

در مورد شبکه‌های عصبی عمیقی که در نورون‌های آن‌ها از تابع فعالیت پله‌ای استفاده شده باشد، با توجه به اینکه هر یک از این نورون‌ها در فضای ورودی خود یک ابرصفحه ایجاد می‌نمایند، قادر خواهند بود که نمونه‌های ورودی را در دو طبقه متمایز در خروجی آن نورون تفکیک نمایند. چنین شبکه‌هایی در یادگیری نمونه‌ها، برای اینکه قادر باشند تمامی نمونه‌های متمایز را در خروجی به‌صورت متفاوت بیان نمایند، لازمه‌اش آن است که در همه لایه‌ها، تمامی نمونه‌های غیرهم‌طبقه، با کدهای متمایزی بیان شده باشند. برای این‌منظور ضروریست که در فضای ورودی، مابین هر دو نمونه غیرهم‌طبقه، حداقل یک ابرصفحه عبور نماید و همین وضعیت در تمامی لایه‌های بعدی شبکه نیز ادامه یابد. به این معنی که در تمامی لایه‌ها بایستی به‌ازای هر طبقه یک ناحیه مجزا توسط ابرصفحات آن لایه شکل گرفته باشد تا در هیچ لایه‌ای دو نمونه مجزا با یک بیان واحد ارائه نشوند و کاهش بعد در اطلاعات متمایزگر صورت نگیرد. به بیان دیگر هنگام تعلیم شبکه‌های عصبی عمیق توسط الگوریتم پس‌انتشار خطا، کمینه‌های موضعی که در مسیر همگرایی یادگیری شبکه مانع ایجاد می‌کنند، غالباً ناشی از قرار نگرفتن ابرصفحات نورون‌ها در لایه‌های

داده می‌شوند. برخلاف BNN‌های مسیر جلو، در مسیر عقب‌سو، وزن‌های خروجی BNN‌ها در ساختار شبکه عمیق استفاده خواهند شد.

لایه گلوگاه آن‌ها آموزش داده می‌شوند. در مسیر عقب‌سو نیز BNN اول با بردار خروجی و BNN‌های بعدی با تصویر بردار ورودی BNN قبلی در لایه گلوگاه آن آموزش



شکل (۱): نحوه تجزیه یک شبکه عصبی عمیق در روش پیش‌تعلیم لایه‌به‌لایه دوسویه

لایه خروجی تعریف شده‌اند. در این روابط Y_{j-1}^B و Y_{i-1}^F تصاویر ورودی BNN‌های $(i-1)$ و $(j-1)$ در دو مسیر مطابق روابط (۳) و (۴) در لایه گلوگاه آن‌ها می‌باشند.

$$Y_{i-1}^F = f(Y_{i-2}^F W_{i-1} - b) \quad n \geq i \geq 2 \quad (3)$$

$$Y_{j-1}^B = f(Y_j^B V_i - b) \quad 2n \geq i \geq n+2 \quad (4)$$

همچنین Y_0^F بردار ورودی BNN اول در مسیر جلو، برابر با X همان بردار ورودی شبکه عصبی عمیق و Y_{2n}^B بردار ورودی BNN اول در مسیر عقب‌سو، برابر با بردار L خروجی مطلوب شبکه عصبی عمیق می‌باشند. Z_j^B و Z_j^F نیز که خروجی‌های BNN i ام و j ام در مسیر جلو

برای پیش‌تعلیم با روش لایه‌به‌لایه دوسویه برای تعیین مقادیر وزن‌ها در مسیر جلو، توابع هزینه محلی بر مبنای بازسازی صحیح ورودی شبکه عصبی عمیق و در مسیر عقب‌سو مبتنی بر بازسازی خروجی مطلوب آن تعریف می‌شوند.

$$E_i^{LF} = \frac{1}{2} \|Y_{i-1}^F - Z_i^F\|^2 \quad n-1 \geq i \geq 1 \quad (1)$$

$$E_j^{LB} = \frac{1}{2} \|Y_{j-1}^B - Z_j^B\|^2 \quad 2n+1 \geq j \geq n+1 \quad (2)$$

در رابطه‌های (۱) و (۲) به ترتیب خطای خودانجمنی BNN‌های i ام و j ام در مسیر جلو و عقب‌سو می‌باشند که برابر با خطای بازسازی ورودی در

و عقب‌سو می‌باشند، در روابط (۵) و (۶) تعریف شده‌اند.

$$Z_i^F = f(Y_i^F V_i - b) \quad n-1 \geq i \geq 1 \quad (5)$$

$$Z_j^B = f(Y_{j-1}^B W_j - b) \quad 2n+1 \geq j \geq n+1 \quad (6)$$

در روابط فوق b مقدار سطح آستانه را برای تابع فعالیت نورون‌ها تعیین می‌کند. این روند تا وزن‌های $h-1$ از مسیر جلو سو و وزن $h+2$ از مسیر عقب‌سو ادامه می‌یابد. سپس تصویر ورودی‌ها در لایه گلوگاه دو BNN آخر از دو مسیر محاسبه و جهت تعلیم یک شبکه دیگرانجمنی یک لایه پنهان متشکل از وزن‌های W_n و W_{n+1} مورد استفاده قرار می‌گیرند. بدین صورت مقادیر اولیه تمام وزن‌های شبکه عمیق به دست می‌آید که تعلیم آن‌ها مبتنی بر الگوریتم پس‌انتشار خطا صورت می‌گیرد.

بعد از انجام مراحل پیش‌تعلیم، وزن‌های به دست آمده به عنوان وزن‌های اولیه در ساختار شبکه عصبی عمیق یکپارچه منظور شده و به کمک عملیات پس‌انتشار خطا مقادیر دقیق‌تر ماتریس‌های وزن شبکه بدست می‌آید.

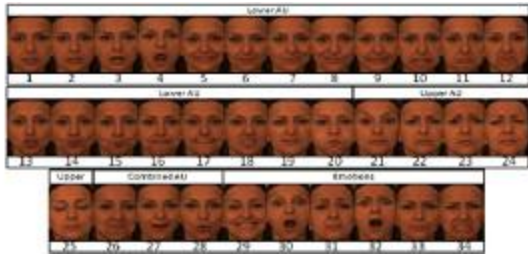
۳- پیاده‌سازی‌ها و نتایج

در این بخش به ارزیابی روش پیش‌تعلیم لایه‌به‌لایه دوسویه، به منظور تنظیم اولیه مناسب وزن‌های شبکه‌های عصبی عمیق دیگرانجمنی، پرداخته می‌شود. بدین منظور کارایی آن در تعلیم شبکه‌های عصبی طبقه‌بندی‌کننده عمیق با هدف بازشناسی در چند پایگاه داده مختلف بررسی می‌گردد. لذا در ادامه سه پایگاه داده مورد استفاده معرفی می‌شوند.

۳-۱- پایگاه داده‌ها

۳-۱-۱- پایگاه داده بسفروس^۴

پایگاه داده بسفروس [۱۶] شامل تصاویر سه بعدی و دویعدی از چهره ۱۰۵ سوژه می‌باشد که در دانشگاه بگازرسی^۵ ترکیه جمع‌آوری شده است. این پایگاه داده شامل مجموعه‌ای غنی از حالات چهره، چرخش سر و انواع مختلف انسداد می‌باشد که مجموعه مربوط به حالات مختلف چهره برای ارزیابی در این مقاله مورد استفاده قرار گرفته است. شکل (۲) این حالت‌ها را برای یکی از افراد



شکل (۲): تصاویر مربوط به حالت‌های مختلف چهره برای یکی از افراد از پایگاه داده بسفروس [۱۶].

تصاویر ۹۵ نفر از این مجموعه به طور تصادفی جهت بررسی انتخاب شد. این تصاویر به صورت سیاه و سفید با ۲۵۶ سطح خاکستری مورد استفاده قرار گرفت. همچنین جهت کاهش بار محاسبات وضوح این تصاویر به $114 * 92$ کاهش داده شد.

در این تحقیق، این پایگاه داده برای بازشناسی چهره استفاده شده است. از آنجا که تعداد تصاویر مربوط به حالات مختلف چهره در آن برای تمام افراد یکسان نمی‌باشد، تصاویر مربوط به ۱۲ حالت مشترک بین تمام افراد برای آموزش و سایر تصاویر برای آزمون کنار گذاشته شد. در نهایت مجموعه‌های آموزش و آزمون به ترتیب شامل ۱۱۲۶ و ۱۴۴۳ تصویر شدند. برچسب افراد نیز به صورت کدهای دودویی ۷تایی که بیانگر ۹۵ کد مختلف معرف هویت افراد می‌باشند، تعریف شد.

۳-۱-۲- پایگاه داده CK+

پایگاه داده CK+ توسعه‌یافته پایگاه داده کوهن-کند^۶ می‌باشد [۱۷]. در پایگاه داده کوهن-کند دنباله‌ای از تصاویر برای ۶ حالت احساسی افراد موجود است که از حالت خنثی تا حالت موردنظر تغییر می‌کنند.

این حالت‌های احساسی شامل حالت‌های شاد، متعجب، خشمگین، ترسیده، متغیر و ناراحت می‌باشند. همچنین برای هر تصویر چهره، برداری از نشانه‌ها وجود دارد که می‌توان از آن برای تراز کردن چهره‌ها استفاده نمود. در این مقاله چهره‌ها طوری تراز شده‌اند که چشم‌ها در یک راستای افقی قرار بگیرند. اطلاعات اضافه اطراف تصاویر نیز به نحوی

ورودی شبکه، این آرایه باید به‌صورت یک بعدی ارائه شود، هر تصویر به یک بردار $MN \times 1$ تبدیل می‌شود. همچنین مقادیر پیکسل‌های هر تصویر چهره به محدوده صفر تا یک بهنجار می‌شوند.

۳-۳- مدل

ساختار شبکه عصبی طبقه‌بندی‌کننده عمیق، جهت بازشناسی در سه پایگاه داده معرفی شده، در جدول (۱) آورده شده است. این شبکه یک شبکه با سه لایه پنهان می‌باشد که ورودی آن مقادیر روشنایی پیکسل‌های تصاویر و خروجی آن برچسب‌های تعریف شده متناظر می‌باشند. این شبکه برای هر مجموعه داده به‌صورت جداگانه تعلیم داده می‌شود و نتایج آن ارزیابی می‌گردد. برای نشان دادن کارایی روش پیش‌تعلیم لایه‌به‌لایه دوسویه، ابتدا مقادیر اولیه تصادفی برای وزن‌ها قرار داده می‌شود و تعلیم صورت می‌گیرد. در مرحله بعد، با استفاده از روش پیش‌تعلیم پیشنهادی، مقادیر اولیه وزن‌ها تعیین و سپس شبکه تعلیم داده می‌شود.

جدول (۱): پارامترهای شبکه عصبی عمیق بازشناس^۷

(DCNN)	
۳	تعداد لایه‌های پنهان
۵۰-۲۰۰-۴۰۰	تعداد نوروں لایه پنهان
غیرخطی تک قطبی	تابع نوروں‌های لایه پنهان
غیرخطی تک قطبی	تابع نوروں‌های لایه خروجی
۰/۰۰۱	ضریب یادگیری
۰/۷	ضریب گشتاور

۴-۳- روش پیش‌تعلیم لایه‌به‌لایه دوسویه برای

پیش‌تعلیم شبکه عمیق بازشناس

مطابق روش ارائه شده در بخش (۲)، این شبکه عمیق به دو BNN تک لایه و یک شبکه دیگرانجمنی یک لایه پنهان شکسته می‌شود. یکی از BNNها با بردار پیکسل‌ها و دیگری با بردار کدهای دودویی که به‌عنوان برچسب‌های متناظر تعریف شده‌اند، به‌طور خودانجمنی تعلیم داده

حذف شده است که همه جزئیات مهم در تصویر موجود باشند. در نهایت ابعاد همه تصاویر به 50×50 تبدیل شدند. در شکل (۳) نمونه‌هایی از تصاویر این پایگاه داده برای حالت‌های احساسی مورد استفاده آورده شده‌اند.



شکل (۳): نمونه‌هایی از تصاویر پایگاه داده کوهن-کند برای

حالت‌های احساسی مختلف بعد از تراز نمودن چهره‌ها و حذف اطلاعات اضافه اطراف تصاویر

همچنین تصاویر نیمه اول هر یک از این دسته‌ها حذف شدند، تا تصاویری که واقعاً نمایانگر حالت احساسی مربوطه هستند، نگه داشته شوند. در این مقاله، این پایگاه داده برای بازشناسی حالت‌های احساسی استفاده شده است. لذا تصاویر به دو دسته تقسیم شده‌اند. مجموعه آموزش، تصاویر مربوط به ۹۶ نفر (۲۸۱۶ تصویر) می‌باشد که جهت آموزش شبکه عصبی عمیق بازشناس استفاده می‌شود. تصاویر مربوط به ۱۰ نفر باقیمانده (۳۵۱ تصویر) نیز برای آزمون به‌کار می‌روند. برچسب حالت‌های احساسی نیز به‌صورت کدهای دودویی ۴تایی که معرف حالت چهره افراد می‌باشند، تعریف شده‌اند.

۳-۱-۳- پایگاه MNIST

این پایگاه شامل ارقام دستنوشته 0 تا 9 انگلیسی می‌باشد که تصویر هر رقم آن به‌صورت یک تصویر خاکستری 28×28 پیکسل ذخیره شده است. مجموعه آموزش آن شامل ۶۰۰۰۰ تصویر و مجموعه آزمون حاوی ۱۰۰۰۰ تصویر می‌باشد [۱۸]. برچسب‌های تعریف شده در این مقاله نیز کدهای دودویی ۳تایی بیان‌کننده عدد می‌باشند.

۲-۳- پیش‌پردازش

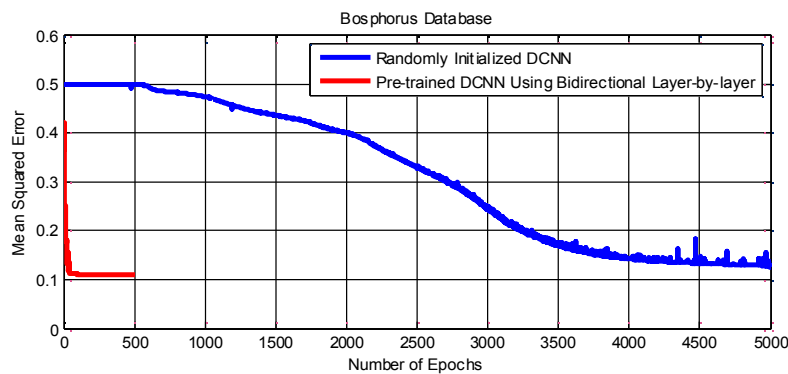
هر تصویر به‌صورت یک آرایه دو بعدی $M \times N$ از مقادیر روشنایی در نظر گرفته می‌شود. اما از آنجاکه در

می‌شوند. تعداد تکرارهای تعلیم نیز در هر مرحله پیش‌تعلیم، برای دادگان چهره شامل تصاویر پایگاه داده‌های بسفروس و CK+ ۵۰۰ تکرار و برای دادگان دستنوشتر با توجه به حجم بالای دادگان، ۱۰ تکرار فرض شده است. چراکه تعلیم بیش از اندازه در مراحل پیش‌تعلیم منجر به تنظیم بیش از حد وزن‌ها روی دادگان آموزش می‌گردد که در نتیجه آن قدرت تعمیم شبکه عمیق پایین خواهد آمد. سپس تصویر ورودی‌ها در لایه پنهان هر BNN محاسبه و جهت تعلیم شبکه دیگرانجمنی یک لایه پنهان مورد استفاده قرار می‌گیرند. بدین‌صورت مقادیر اولیه تمام وزن‌های شبکه

عمیق به‌دست می‌آیند.

۳-۵- نتایج

نمودار شکل (۴)، تغییرات خطای تعلیم را برای دو شبکه عمیق بازشناس فرد، یکی با مقداردهی اولیه تصادفی و دیگری با مقداردهی اولیه توسط روش پیش‌تعلیم پیشنهادی، روی پایگاه داده بسفروس نشان می‌دهد. همان‌طور که مشاهده می‌شود با تنظیم اولیه مناسب وزن‌های شبکه به‌جای قرار دادن مقادیر تصادفی برای آن‌ها، سرعت همگرایی تعلیم به‌طور چشمگیری افزایش می‌یابد.



شکل (۴): نمودار تغییرات خطای تعلیم شبکه عمیق بازشناس فرد با دو روش مقداردهی اولیه وزن‌ها به‌صورت تصادفی و با استفاده از پیش‌تعلیم لایه‌به‌لایه دوسویه روی پایگاه داده بسفروس

به‌گونه‌ایکه در همان تکرارهای اولیه به مقدار کمینه خود می‌رسد. برای شبکه با مقداردهی تصادفی (با توجه به روند کاهشی خطای تعلیم) وزن‌های آخرین تکرار تعلیم و برای شبکه پیش‌تعلیم شده وزن‌های تکرار ۴۰ام (بعد از آن خطای تعلیم ثابت مانده است.) برای بازشناسی افراد در داده آزمون به‌کار رفت.

نتایج جدول (۲) نشان می‌دهد که با استفاده از این روش پیش‌تعلیم، نتایج بازشناسی فرد برای دادگان آزمون حدود ۱۱٪ بهبود یافته است. قابل توجه است که نتیجه حاصل از شبکه پیش‌تعلیم شده مربوط به تکرار ۴۰ام تعلیم می‌باشد که با لحاظ نمودن مراحل پیش‌تعلیم که شامل تعلیم ۳ شبکه یک لایه پنهان با ۵۰۰ تکرار تعلیم برای هر یک می‌باشد، در مقایسه با ۵۰۰۰ تکرار تعلیم شبکه مقداردهی شده تصادفی مقدار کمتری است. خصوصاً اینکه در مراحل پیش‌تعلیم شبکه‌های یک لایه پنهان تعلیم داده می‌شوند که مدت زمان تعلیم آن‌ها در هر تکرار در مقایسه با تعلیم شبکه سه لایه پنهان بسیار کمتر می‌باشد.

جدول (۲): مقایسه درصد صحت بازشناسی افراد در پایگاه داده بسفروس توسط شبکه عصبی سه لایه پنهان بعد از تنظیم

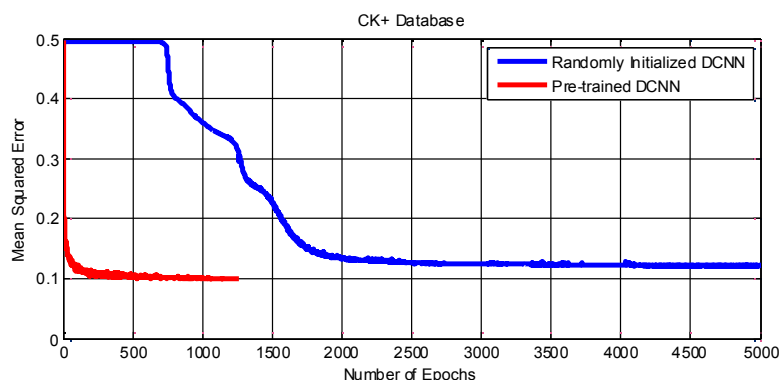
دقیق وزن‌ها

نحوه مقداردهی اولیه وزن‌های شبکه عصبی عمیق	تعداد تکرارها	داده آموزش	داده آزمون
تصادفی	۵۰۰۰	٪۹۹٫۷۲	٪۸۶٫۵۲
پیش‌تعلیم لایه‌به‌لایه دوسویه	۴۰	٪۱۰۰	٪۹۷٫۱۶

در نمودار شکل (۵) و جدول (۳) نتایج روی پایگاه داده CK+ با هدف بازشناسی حالت احساسی در تصاویر چهره آورده شده است.

روش پیش‌تعلیم لایه‌به‌لایه دوسویه برای تعلیم شبکه‌های عصبی عمیق

همان‌طور که در نمودار شکل (۵) مشاهده می‌شود، در نمودار خطای شبکه با مقادیر اولیه تصادفی بعد از تکرار ۲۵۰۰ام تعلیم، کاهش معناداری صورت نمی‌گیرد. لذا وزن‌های این مرحله برای شبکه بازشناس حالت استفاده می‌شود. برای شبکه پیش‌تعلیم شده نیز وزن‌های تکرار ۱۹۰ام تعلیم به‌کار رفته است.



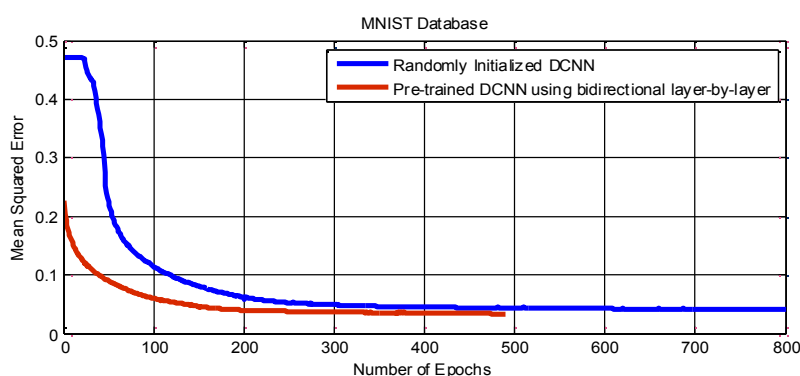
شکل (۵): نمودار تغییرات خطای تعلیم شبکه عمیق بازشناس حالت احساسی چهره با دو روش مقداردهی اولیه وزن‌ها به صورت تصادفی و با استفاده از پیش‌تعلیم لایه‌به‌لایه دوسویه روی پایگاه داده CK+

در مورد پایگاه داده ارقام دست‌نویسار MNIST با توجه به تعداد زیاد نمونه‌ها، تعداد تکرارهای تعلیم در مراحل پیش‌تعلیم ۱۰ تکرار در نظر گرفته شد. آزمایشات نیز نشان داد که بعد از این تعداد، خطا در مراحل پیش‌تعلیم برای این داده بهبود نمی‌یابد. در نمودار شکل (۶) تغییرات خطای تعلیم شبکه عمیق بازشناس ارقام در مرحله تنظیم دقیق وزن‌ها، برای دو روش مقداردهی اولیه وزن‌ها به صورت تصادفی و با استفاده از پیش‌تعلیم لایه‌به‌لایه دوسویه آورده شده است.

نتایج بازشناسی در جدول (۳) نشان می‌دهند که استفاده از روش پیش‌تعلیم لایه‌به‌لایه دوسویه برای مقداردهی اولیه وزن‌های شبکه بازشناس، درصد صحت بازشناسی حالت را ۱۴٪ بهبود می‌دهد.

جدول (۳): مقایسه درصد صحت بازشناسی حالت‌های احساسی توسط شبکه عصبی سه لایه پنهان بعد از تنظیم دقیق وزن‌ها روی پایگاه داده CK+

روش مقداردهی اولیه وزن‌های شبکه عصبی عمیق	تعداد تکرارها	داده آموزش	داده آزمون
تصادفی	۲۵۰۰	٪۹۹،۴۱	٪۷۳،۷۰
پیش‌تعلیم لایه‌به‌لایه دوسویه	۹۰	٪۹۹،۹۶	٪۸۷،۲۹



شکل (۶): نمودار تغییرات خطای تعلیم شبکه عمیق بازشناس ارقام روی پایگاه داده دست‌نویسار با دو روش مقداردهی اولیه وزن‌ها به صورت تصادفی و با استفاده از پیش‌تعلیم لایه‌به‌لایه دوسویه

مقاله روش پیش‌تعلیم لایه‌به‌لایه دوسویه جهت مقداردهی مناسب وزن‌های اولیه شبکه‌های چندلایه با عملکرد دیگرانجمنی ارائه شده است که به‌طور موازی در یک مسیر دوسویه به‌طور جلوسو و عقب‌سو به پیش‌تعلیم وزن‌های شبکه می‌پردازد. این روش برخلاف روش‌های پیش‌تعلیم پیشین [۹،۱۴] که صرفاً برای شبکه‌های عصبی خودانجمنی تعریف شده بودند، برای کاربردهای دیگرانجمنی شبکه‌های عصبی عمیق ارائه شده است.

این روش برای پیش‌تعلیم وزن‌های سه شبکه عصبی عمیق بازناس فرد، طبقه‌بندی‌کننده حالت‌های احساسی و بازناس ارقام دستنوشته مورد استفاده قرار گرفت و نشان داده شد که با به‌کارگیری این روش پیش‌تعلیم سرعت همگرایی تعلیم به‌طور چشمگیری افزایش می‌یابد. همچنین میزان بازناسی در پایگاه داده‌های چهره به میزان قابل توجهی بهبود می‌یابد که حاکی از افزایش قدرت تعمیم شبکه با استفاده از این روش پیش‌تعلیم می‌باشد.

مراجع

- [1] Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U., and Poggio, T., "A quantitative theory of immediate visual recognition", *Progress in Brain Research, Computational Neuroscience: Theoretical Insights into Brain Function*, Vol. 165, pp. 33–56, 2007.
- [2] Erhan, D., Manzagol, P.A. Bengio, Y. Bengio, S. Vincent, P. "The difficulty of training deep architectures and the effect of unsupervised pre-training", In *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics (AISTATS'09)*, pp. 153-160, 2009.
- [3] Bengio, Y., "Learning deep architectures for AI", *Foundations and Trends® in Machine Learning*, Vol 2, No. 1, pp. 1-127, 2009.
- [4] Erhan, D., "Understanding deep architectures and the effect of unsupervised pre-training", Ph.D. Thesis, Université de Montréal, 2011.
- [5] Bengio, Y., "Evolving culture vs local minima", arXiv preprint arXiv: 1203.2990, 2012.
- [6] Plath, N., "Extracting low-dimensional features by means of deep network architectures", Ph.D. Thesis, Technische Universität Berlin, April 2008.
- [7] Seyyedsalehi S.Z. and Seyyedsalehi S.A., "Simultaneous learning of nonlinear manifolds based on the bottleneck neural

در جدول (۴) خطای بازناسی ارقام برای دو روش مقداردهی اولیه وزن‌های شبکه مشاهده می‌شود. برای دادگان دستنوشته بهبودی در بازناسی ارقام توسط شبکه پیش‌تعلیم شده حاصل نشده است. می‌توان عنوان نمود که مشکل کمینه‌های موضعی در روند یادگیری شبکه‌های عمیق در مسائل پیچیده نظیر چهره و حالت چهره بروز جدی‌تری پدید می‌کند، لذا استفاده از پیش‌تعلیم مناسب در روند تعلیم، کارایی خود را بیشتر نشان می‌دهد. اما برای ارقام دستنوشته با توجه به ساده‌تر بودن تکلیف یادگیری، شبکه مقداردهی شده با وزن‌های اولیه تصادفی، نتیجه مطلوبی دارد به‌گونه‌ایکه پیش‌تعلیم در نتایج نتوانسته بهبود معناداری ایجاد نماید. البته باید عنوان نمود که نتیجه حاصل از شبکه با پیش‌تعلیم سه برابر سریعتر حاصل شده است. خصوصاً اینکه مراحل پیش‌تعلیم آن کوتاه بوده است.

جدول (۴): مقایسه درصد صحت بازناسی ارقام دستنوشته توسط شبکه عصبی سه لایه پنهان بعد از تنظیم دقیق وزن‌ها

روش‌های مقداردهی اولیه وزن‌های شبکه عصبی عمیق	تعداد تکرارها	داده آموزش	داده آزمون
تصادفی	۶۰۰	٪۹۹٫۸۸	٪۹۸٫۲۴
پیش‌تعلیم لایه‌به‌لایه دوسویه	۲۰۰	٪۹۹٫۹۲	٪۹۸٫۵۶

۴- نتیجه‌گیری

در برخی موارد شبکه‌هایی با یک یا دو لایه پنهان در نگاشت داده ورودی به فضای خروجی موفق عمل نمی‌کنند. چراکه برای داده‌های پیچیده‌تر با بعد بالا مانند تصاویر چهره تبدیل ناگهانی ویژگی‌های داده ورودی به خروجی که می‌تواند برچسب طبقات ورودی باشند، خطای شبکه را به میزان قابل توجهی افزایش می‌دهد. لذا نیاز به ساختارهای عمیق‌تر شبکه‌های عصبی می‌باشد تا در طی لایه‌های متعدد مؤلفه‌های عمیق‌تر داده‌ها استخراج شوند و سپس به کمک آن‌ها نگاشت مناسب به خروجی صورت گیرد. این در حالی است که با افزایش تعداد لایه‌های شبکه‌های عصبی، همگرایی تعلیم آن‌ها به دلیل افزایش کمینه‌های موضعی با مشکل مواجه می‌شود. به‌منظور حل این مشکل تاکنون برخی روش‌ها برای پیش‌تعلیم ساختارهای عمیق شبکه‌های عصبی در کاربردهای خودانجمنی ارائه شده‌اند. اما در این

- [13] Ghasemi M., "Nonlinear independent component analysis of Speech signal", M.S. Thesis, Amirkabir University of Technology, 2006.
- [14] Nejadgholi I., "A brain-inspired model of feature extraction and binding considering their interactions", Ph.D. Thesis, Amirkabir University of Technology, 2012.
- [15] Hinton, G.E. and Salakhutdinov, R.R., "Reducing the dimensionality of data with neural networks", *Science*, Vol. 313, No. 5786, pp. 504–507, 2006.
- [16] Savran, A., Alyüz, N., Dibeklioglu, H., Çeliktutan, O., Gökberk, B., Sankur, B., Akarun, L., "Bosphorus Database for 3D Face Analysis", BIOID 2008, Roskilde University, Denmark, May 2008.
- [17] Lucey, P., et al. "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression", in *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2010 IEEE Computer Society Conference on. 2010. IEEE.
- [18] LeCun, Y. and Cortes, C., "The MNIST database of handwritten digits", 1998.
- network", *Neural Processing Letters*, Vol. 40, No. 2, pp. 191-209. 2014.
- [8] Seyyedsalehi S.Z. and Seyyedsalehi S.A., "New fast pre-training method for training of deep neural network", *International Journal of Signal and Data Processing*, Vol. 10, No. 1, pp. 13–26, 2013.
- [9] Seyyedsalehi S.Z. and Seyyedsalehi S.A., "New fast pre-training method for training of deep neural network learning", in *Proc. 19th ICBME*, 2012.
- [10] Bengio, Y., Lamblin, P., Popovici, D., and Larochelle, H., "Greedy layer-wise training of deep networks", *Advances in neural information processing systems 19 (NIPS'06)*, pp. 153-160, MIT Press, 2007.
- [11] Erhan, D., Bengio, Y., Courville, A., Manzagol, P.A., Vincent, P., Bengio, S., "Why does unsupervised pre-training help deep learning?", *The Journal of Machine Learning Research*, Vol. 11, pp. 625-660, 2010.
- [12] Nejadgholi I., Seyyedsalehi S.A., "Experiments towards bidirectional neural networks", Technical Report, Research Center of Intelligent Signal Processing, 2004.

¹ Bottleneck Neural Network

² Hinton

³ Restricted Boltzmann Machine

⁴ Bosphorus

⁵ Bogazici

⁶ Cohn-Kanade

⁷ Deep Classifier Neural Network