روش پیش‌تعلیم سریع بر مبنای کمینه‌سازی خطا برای همگرایی یادگیری شبکه‌های عصبی با ساختار عمیق

سیّده زهره سیّدصالحی و سیّد علی سیّدصالحی
دانشکدهٔ مهندسی پزشکی، دانشگاه صنعتی امیرکبیر، تهران، ایران

چکیده:

در این مقاله با توسعهٔ روش‌های موجود و بر مبنای کمینه‌سازی خطا و حفظ تمایز بیشینه بین نمونه‌ها، یک روش پیش‌تعلیم لایه‌به‌لایهٔ سریع و کارا جهت مقداردهی اولیه مناسب وزن‌ها در شبکه‌های عصبی با ساختارهای عمیق ارائه شده است. تعلیم شبکه‌های عصبی عمیق به‌دلیل مواجهه با تعداد بالای کمینه‌های موضعی اغلب همگرا نمی‌شود. درحالی‌که با مقداردهی اولیه مناسب وزن‌های شبکه به‌جای مقادیر تصادفی در ابتدای مسیر تعلیم، می‌توان از بسیاری از کمینه‌های موضعی اجتناب کرد. در این روش شبکهٔ عصبی چندلایه به تعداد متناظری شبکه با یک لایهٔ پنهان شکسته می‌شود و ابتدا این شبکه‌های یک لایهٔ پنهان تعلیم داده می‌شوند. سپس مقادیر وزن حاصل از تعلیم اینها در شبکهٔ عصبی اصلی قرار داده می‌شود و برای تنظیم دقیق وزن‌ها، تعلیم یک‌پارچه صورت می‌گیرد. روش پیشنهادی برای پیش‌تعلیم شبکهٔ عصبی خودانجمنی پنج لایهٔ پنهان جهت استخراج مؤلّفه‌های اساسی غیرخطی چهره برای دادگان بسفروس مورد استفاده قرار گرفت. مقایسهٔ میانگین نتایج شبکه‌های عصبی با مقداردهی اولیه تصادفی و مقداردهی با روش پیش‌تعلیم لایه‌به‌لایه نشان می‌دهد که این روش پیش‌تعلیم، علاوه‌بر اینکه سرعت همگرایی تعلیم را بهبود می‌دهد، قدرت تعمیم شبکه را نیز بالا می‌برد. به‌گونه‌ای‌که با وجود خطای تعلیم یکسان، با به‌کارگیری روش پیش‌تعلیم لایه‌به‌لایه برای مقداردهی اولیه وزن‌ها، خطای بازسازی هر پیکسل ٪۱۳/۶۹ کاهش و درصد صحت بازشناسی تصاویر با استفاده از مؤلّفه‌های استخراج شده حدود ٪۱۰ بهبود داشته است. همچنین بررسی‌ها نشان داد که روش پیش‌تعلیم لایه‌به‌لایه در مقایسه با دو روش پیش‌تعلیم مشرف به هدف و تجزیه به ماشین‌های بولتزمان کارایی بالاتری دارد.

واژگان کلیدی: استخراج مؤلّفه، پیش‌تعلیم، شبکه‌های عصبی، ساختار عمیق، همگرایی .

۱. مقدّمه

توجه به مفاهیم موجود در داده‌ها و ارتباطات و تعاملات مابین آنها نقش مهمی را در کیفیت یادگیری و تعمیم‌دهی شبکه‌های عصبی ایفا می‌کند. این مفاهیم را می‌توان تحت عنوان مؤلّفه‌های دانش مورد یادگیری مدل‌ها و نحوهٔ پردازش شناختی اطّلاعات توسط آنها در نظر گرفت (سیدصالحی، ۱۳۸۳).

سامانه‌های ادراکی اغلب دارای ساختارهایی پیچیده با ابعاد بالا هستند و چکیده‌های مفهومی و به‌عبارتی همان ویژگی‌ها یا مؤلّفه‌ها در یک مجموعهٔ پیچیده از روابط بین این ویژگی‌ها در فضای با ابعاد بالا، به‌صورت نهفته قرار دارند (Seow, 2006). تاکنون روش‌های مختلفی برای استخراج این مؤلّفه‌ها مطرح شده‌اند. دسته‌ای از این روش‌ها، به‌صورت خطی عمل می‌کنند که برای داده‌های با روابط پیچیده و غیرخطی جواب‌گو نمی‌باشند. روش‌های غیرخطی هم که در طی چند سال اخیر ارائه شده‌اند، روی داده‌های مصنوعی غیرخطی نتایج خوبی ارائه داده‌اند ولی برای داده‌های واقعی حتی در مقایسه با برخی روش‌های خطی کارایی پایین‌تری دارند (Maaten et al., 2008). یکی از رویکردها در این زمینه به‌کارگیری شبکه‌های عصبی

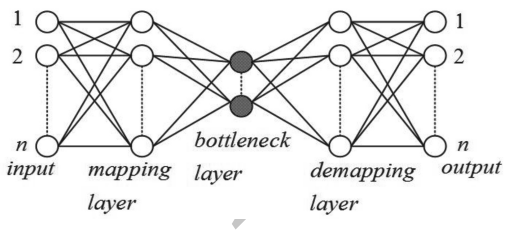
Archive of SID
www.SID.ir

مصنوعی می‌باشد. این مدل‌ها با توجه به اینکه دارای توجیه زیستی هستند، به‌نظر می‌رسد که استخراج مؤلّفه‌ها از طریق آنها، هم‌خوانی بیشتری با نحوهٔ عملکرد مغز انسان در ادراک و شناخت داشته باشد. در مغز انسان سیگنال ورودی مورد ادراک، در چندین مرحله بازنمایی می‌شود. هر مرحله مربوط به یک ناحیه متفاوت از قشر مخ است. همچنین مغز پردازش اطّلاعات را از طریق چندین مرحله بازنمایی و تبدیل انجام می‌دهد (Serre *et al.*, 2007). با توجه به وجود ساختارهای عمیق در مغز انسان، محققان به سمت تعلیم شبکه‌های عصبی چندلایهٔ عمیق گرایش پیدا کرده‌اند. در این ساختارها امکان استخراج سلسله‌مراتبی مؤلّفه‌ها فراهم می‌شود. به‌گونه‌ای‌که مؤلّفه‌های سطح بالا به‌صورت ترکیبی از مؤلّفه‌های سطح پایین در چندین سطح تشکیل می‌گردند (Bengio, 2009). اما تعلیم معمول شبکه‌های با بیش از سه لایه اغلب به نتایج ضعیفی منجر می‌شود. چراکه علاوه بر زمان‌بر بودن تعلیم ساختارهای عمیق شبکه، با افزایش تعداد لایه‌ها و نورون‌ها، مشکل کمینه‌های موضعی محسوس‌تر می‌شود (Plath, 2008). به‌گونه‌ای‌که در اغلب موارد، تعلیم همگرا نمی‌شود. بررسی‌ها نشان می‌دهد که در این ساختارها تعداد کمینه‌های موضعی فعال که به نوع توابع و مقداردهی اولیه پارامترهای شبکه بستگی دارند، زیاد است (Bengio, 2012). لذا مقداردهی اولیه مناسب وزن‌های شبکه جهت همگرا کردن تعلیم آن ضروری به‌نظر می‌رسد.

در این مقاله با استفاده از تجارب موجود قبلی بر مبنای روش مشرف به هدف (نژادقلی،۱۳۸۳،۱۳۹۱) و روش پیش‌تعلیم ارائه‌شده توسط هینتن[1] با استفاده از شبکه‌های بولتزمان (Hinton & Salakhutdinov, 2006) و با توسعه روی آنها، برمبنای تحلیل کمینه‌سازی خطای جمعی و حفظ تمایز بیشینه بین نمونه‌ها به‌صورت مرحله به مرحله، روشی برای تعیین مقادیر اولیه وزن‌های شبکه ارائه شده است. این روش به‌عنوان یک روش پیش‌تعلیم بدون سرپرستی با تغییر شرایط اولیه، الگوریتم پس‌انتشار خطا را در جهت رسیدن به کمینهٔ مطلوب کمک می‌کند. روش ارائه‌شده در تعلیم شبکه‌های عصبی عمیق به‌منظور استخراج مؤلّفه‌های اساسی غیرخطی چهره مورد استفاده قرار گرفته است؛ به همین منظور ابتدا در بخش دوم به شرح نحوهٔ استخراج مؤلّفه با استفاده از شبکه‌های عصبی پرداخته، سپس در بخش سوم مشکلات تعلیم ساختارهای عمیق شبکه‌های عصبی و راه‌حل‌های موجود برای آن توضیح داده می‌شود. روش پیشنهادی این مقاله در بخش چهارم شرح داده شده است. بخش پنجم هم به شرح پیاده‌سازی‌ها و نتایج آنها پرداخته و در نهایت جمع‌بندی و نتیجه‌گیری در بخش ششم شرح داده می‌شود.

۲. تحلیل مؤلّفه‌های اساسی غیرخطی[2] با استفاده از شبکه‌های عصبی

یکی از روش‌های پیاده‌سازی تحلیل مؤلّفه‌های اساسی غیرخطی استفاده از شبکه‌های عصبی خودانجمنی است. شبکهٔ خودانجمنی که در سال ۱۹۹۱ توسط *کرامر*[3] معرفی شد، به‌عنوان شبکهٔ تحلیل‌گر مؤلّفه‌های اساسی غیرخطی شناخته می‌شود. به‌دلیل ساختار ویژه‌ای که این شبکه دارد، اغلب به آن شبکه گلوگاه (BNN)[4] نیز می‌گویند (Kramer, 1991). در این ساختار رابطهٔ بین مجموعه داده‌های ورودی و بردار لایهٔ پنهان به‌گونه‌ای مدل می‌شود که خطای بین خروجی واقعی و بردار هدف که برابر ورودی است، کمینه شود. شبکه‌های جلوسوی چندلایه در یک روش یادگیری باسرپرستی، درعمل نقش استخراج بدون سرپرستی ساختار داده‌ها را می‌توانند ایفا کنند. یک شبکهٔ BNN بنا به نوع تابع فعالیت استفاده شده در هر لایه، هر دو نوع رابطه خطی و غیرخطی بین متغیرها را می‌تواند پیاده‌سازی کند.



(شکل۱): یک شبکه گلوگاه با سه لایه پنهان
(Daszykowski et al., 2003).

در شکل (۱) نمونه‌ای BNN آورده شده است. برای نگاشت از فضای دادگان n بُعدی به فضای لایهٔ پنهان با ابعاد کمتر،

[1] Hinton

[2] Nonlinear Principal Component Analysis
[3] Kramer
[4] Bottleneck Neural Network

Archive of SID
www.SID.ir

BNN حداقل سه لایه دارد که به‌ترتیب لایهٔ نگاشت[1]، لایهٔ گلوگاه و لایهٔ نگاشت معکوس[2] نامیده می‌شود. تعداد نورون‌های بیشتر در لایهٔ نگاشت خصوصیات غیرخطی بیشتری را می‌تواند مدل کند. در حین تعلیم، داده‌های ورودی به تعداد متغیرهای کمتری در لایه گلوگاه فشرده می‌شوند. تعداد این متغیرها بستگی به تعداد نورون‌های لایه گلوگاه دارد که باید از قبل مشخص باشند و در حین تعلیم ثابت است.

تعداد گره‌های لایه ورودی و نورون‌های لایه خروجی برابر و مساوی بعد ورودی است. نقش شبکه، از لایه ورودی تا لایه گلوگاه استخراج مؤلّفه و کاهش بعد است. از لایهٔ گلوگاه به سمت خروجی نیز بازسازی داده‌ها از روی مؤلّفه‌های استخراج‌شده صورت می‌گیرد. در این شبکه وزن‌ها با الگوریتم پس‌انتشار خطا بهینه می‌شوند (Daszykowski et al., 2003).

در اغلب موارد شبکهٔ یک لایهٔ پنهان در استخراج مؤلّفه‌های دادهٔ ورودی کارا نیست. درعمل برای داده‌های پیچیده‌تر با بُعد بالا مانند تصاویر چهره تبدیل ناگهانی ویژگی‌های داده ورودی به تعداد مؤلفّهٔ محدود مورد انتظار حاصل از یک لایه، کیفیت بازسازی را بسیار مخدوش کرده و خطای شبکه را به میزان قابل توجهی افزایش می‌دهد. هینتن عنوان می‌کند که به‌لحاظ تجربی وجود یک کاهش ملایم در تعداد نورون‌های لایه‌ها برای رسیدن به مؤلّفه‌های نهایی محدود، ضروری است (Hinton & Salakhutdinov, 2006). لذا به‌نظر می‌رسد نیاز به ساختارهای عمیق BNN است که در این مقاله شبکه گلوگاه عمیق[3] (DBNN) نامیده شده است.

۳. تعلیم ساختارهای عمیق شبکه‌های عصبی

مرور برخی نتایج در زمینه یادگیری ماشین نشان می‌دهد که تعلیم ساختارهای عمیق‌تر شبکه‌های عصبی در مقایسه با ساختارهای کم‌عمق مشکل‌تر است. وقتی تلاش می‌شود تمام لایه‌ها از طریق یک تابع معیار مثل میزان شباهت به ورودی‌ها یا شباهت شرطی به طبقه‌های هدف ورودی‌های داده شده، تعلیم داده شوند نتایج ممکن است بدتر از مدل‌های کم‌عمق است. یکی از دلایل مهم آن، وابستگی زیاد بین پارامترهای شبکه در لایه‌های مختلف در حین تعلیم شبکه‌های عمیق است. در طی آموزش شبکه، وزن‌ها باید به‌گونه‌ای بهینه شوند که به‌طور همزمان هر دو شرط زیر فراهم شوند.

۱. تنظیم لایه‌های سطحی‌تر به‌نحوی که ورودی مناسبی را برای تنظیم نهایی (پایان آموزش) لایه‌های عمیق‌تر فراهم کنند.

۲. تنظیم لایه‌های عمیق‌تر به‌نحوی که در پایان آموزش از مولفه‌های استخراج‌شده توسط لایه‌های سطحی استفاده مناسبی نمایند.

تنظیم وزن‌های لایه‌های عمیق‌تر در صورتی‌که وزن‌های لایه‌های سطحی‌تر مشخص باشند، به‌سادگی صورت می‌گیرد. اما آنچه که مسئله را مشکل می‌کند تنظیم همزمان وزن‌های لایه‌های عمیق و سطحی است که باعث می‌شود شیب تابع هدف توسط مقادیر محلی فراهم شده توسط تنظیمات فعلی پارامترهای شبکه، محدود شود. بررسی‌ها نشان داده است که روش استاندارد آموزش، پارامترها را در محدوده‌ای تنظیم می‌کند که منجر به تعمیم پایین شبکه می‌شود (Erhan, 2011). لذا در کنار رویکرد مستقیم و معمول تعلیم یک شبکهٔ عصبی چندلایه که از روش پس‌انتشار خطا استفاده می‌کند، رویکردهای جدیدی وجود دارند که با بالارفتن پیچیدگی داده‌ها و تعداد نورون‌ها، مفید واقع می‌شوند. چرا که تنها مسئلهٔ زمان در تعلیم ساختارهای عمیق شبکه، مشکل‌ساز نیست؛ بلکه مشکل اصلی کمینه‌های موضعی هستند (Plath, 2008). در ادامه دو راهکار موجود برای این مسئله مطرح می‌شود.

۳-۱- تعلیم با برنامه آموزشی[4]

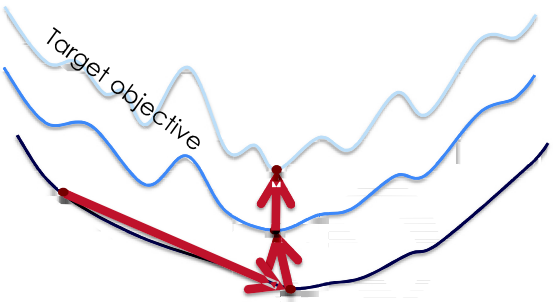
یکی از ایده‌ها برای تعلیم ساختارهای عمیق، تبدیل تدریجی وظیفهٔ یادگیری از یک نسخهٔ ساده به نسخهٔ مشکل آن است. برای مثال اگر هدف، حل یک مسئلهٔ بهینه‌سازی با کمینه‌های موضعی متعدد مطابق شکل (۲) باشد، در ابتدا یک نسخهٔ آسان و هموار، مسئله بهینه‌سازی حل و در ادامه به‌تدریج پیچیده می‌شود.

اساس این ایده به این‌صورت است که یادگیری با جنبه‌های سادهٔ مسئله و یا زیرمسئله‌های آسان آغاز می‌شود و به‌تدریج

[1] Mapping
[2] Damping
[3] Deep Bottleneck Neural Network
[4] Curriculum Learning

Archive of SID
www.SID.ir

در جهت مراحل مشکل‌تر حرکت می‌کند. میزان یادگیری انسان و حیوان نیز در مثال‌های پیچیده هنگامی‌که به همین صورت عمل می‌شود، بالاتر است (Krueger & Dayan, 2009). به‌عبارتی انسان‌ها و حیوانات در یادگیری مثال‌ها اغلب زمانی موفق عمل می‌کنند که مثال‌ها غیرتصادفی و به‌صورت سازماندهی‌شده ارائه شده باشند؛ به‌گونه‌ای‌که بتوانند مفاهیم معنادار را بهتر نشان دهند.



(شکل ۲): تبدیل تدریجی وظیفه یادگیری از یک نسخه ساده به مشکل برای بهینه‌سازی (Bengio et al., 2009).

از دیدگاه بازنمایی داده‌ها این ایده را می‌توان این‌گونه بیان کرد که ابتدا توسط این روش مؤلّفه‌هایی یادگیری می‌شوند که مفاهیم سطح پایین را در خود ذخیره کرده‌اند و سپس از ترکیب اینها برای استخراج مفاهیم سطح بالا در جهت یادگیری ساختارهای پیچیده داده‌های ورودی استفاده می‌شود (Bengio, 2009).

بنجیو[1] و همکارانش با ارائه یک برنامهٔ آموزشی این ایده را ارزیابی کردند. در این روش برای آموزش، ابتدا از مثال‌های ساده‌تر شروع و به‌تدریج سطح مثال‌ها را پیچیده‌تر کردند. درنتیجه ابتدا وزن‌ها برای ساده‌ترین مفاهیم با استفاده از ساده‌ترین مثال‌ها تنظیم و به‌تدریج برای مثال‌های کمی پیچیده‌تر اصلاح شدند و به همین ترتیب ادامه یافت. با توجه به نتایج آزمایش‌ها روی دادگان تصویر و گفتار، این روش سرعت همگرایی شبکه، کارایی و قدرت تعمیم را بهبود می‌دهد (Bengio et al., 2009).

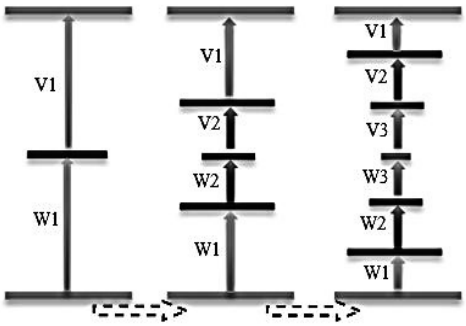
۳-۲- به‌کار گیری روش‌های پیش‌تعلیم

روش دیگر برای تعلیم ساختارهای عمیق، به‌کارگیری روش‌های پیش‌تعلیم است که سعی در یافتن نقطه شروع مناسبی برای وزن‌های شبکه دارند. بررسی‌ها نشان می‌دهد که در تعلیم ساختارهای عمیق، تعلیم مستقیم همهٔ لایه‌ها با هم نه تنها استخراج تمام توان مدل‌سازی را مشکل می‌سازد بلکه با افزایش لایه‌ها درعمل نتایج بدتر هم می‌شود (Bengio et al., 2009; Erhan et al., 2010; Larochelle et al., 2009).

همچنین در آزمایش‌هایی که *ارهان*[2] در سال ۲۰۰۹ انجام داد، نشان داد 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 که در منطقه‌ای از فضای پارامترها که جواب اجازه دارد در آنجا باشد، قرار بگیرند (Erhan et al., 2009).

یکی از روش‌ها جهت پیش‌تعلیم شبکه، روش مشرف به هدف است. این روش اولین‌بار در جهت تعلیم شبکه‌های عصبی دوسویه عمیق در (نژادقلی، ۱۳۸۳) برای استخراج مؤلّفه‌های اساسی استفاده شد و سپس در (قاسمی، ۱۳۸۴) و (نژادقلی، ۱۳۹۱) نیز به‌کار رفت.

این مدل در ابتدا یک شبکهٔ گلوگاه یک‌لایهٔ پنهان است که در طی آموزش، ساختار آن رشد داده می‌شود. شبکهٔ یک‌لایهٔ پنهان ابتدا توسط داده‌های تعلیم، آموزش داده می‌شود تا خطای تعلیم تا حدودی افت نماید. بعد از ثابت‌شدن خطا، دو لایه مطابق شکل (۳)، یکی در بخش نگاشت و دیگری در بخش نگاشت معکوس به ساختار افزوده می‌شود. وزن‌های بین لایهٔ ورودی و لایهٔ جدید و همچنین خروجی و لایهٔ جدید در بخش بازسازی همان وزن‌های قبل و سایر وزن‌ها تصادفی قرار داده می‌شوند. دوباره شبکهٔ آموزش داده می‌شود تا وزن‌های جدید و قبلی تنظیم شوند. این روند رشد ساختار تا رسیدن به تعداد لایهٔ مطلوب ادامه می‌یابد.



(شکل ۳): نحوه رشد عمق شبکه گلوگاه

[1] Bengio

[2] Erhan

Archive of SID
www.SID.ir

روش بعدی، تجزیهٔ شبکه به ماشین‌های بولتزمان است که اولین‌بار توسط هینتن ارائه شد. در این روش در مرحله پیش‌تعلیم به‌منظور تعیین مقادیر اولیه ماتریس‌های وزن، شبکهٔ چندلایه به تعداد متناظری ماشین بولتزمان محدود شده[1] (RBM) شکسته می‌شود. RBMها هریک دارای دو لایهٔ پنهان و آشکار هستند. در ماشین بولتزمان اول با استفاده از داده‌های ورودی و با تعلیم شبکه، مقادیر لایهٔ پنهان و ماتریس وزن تعیین می‌شوند. سپس مقادیر خروجی این لایه برای RBM بعدی به‌عنوان ورودی استفاده می‌شود. این تعلیم لایه‌به‌لایه را می‌توان به هر میزان مورد نیاز تکرار کرد. (Hinton& Salakhutdinov, 2006).

۴. روش پیشنهادی: پیش‌تعلیم لایه‌به‌لایهٔ شبکه‌های عصبی عمیق

شبکه‌های عصبی با ساختار عمیق با استفاده از لایه‌های متعددی از نورون‌ها، ظرفیت یادگیری تبدیلات غیرخطی چندمرحله‌ای و پیچیده را دارند. اما همان‌طورکه عنوان شد با افزایش تعداد لایه‌های شبکه‌های عصبی همگرا کردن تعلیم آنها به‌دلیل وجود کمینه‌های موضعی مشکل و گاهی غیرممکن می‌شود. حال اگر بتوان به‌طریقی تعداد کمینه‌های موضعی ممکن را در مسیر تعلیم شبکه محدود کرد، همگرایی بهبود می‌یابد.

DBNNها که به‌طور معمول برای استخراج مؤلّفه‌ها و فشرده‌سازی غیرخطی داده‌ها مورد استفاده قرار می‌گیرند، به‌دنبال تجزیه غیرخطی اطّلاعات ورودی به مؤلّفه‌های غیرخطی و سپس بازسازی ورودی در خروجی شبکه با کمترین خطا، هستند. اما درصورتی قادر به بازسازی ورودی با دقت مطلوبی می‌باشند که بتوانند در تجزیه و تحلیل اطّلاعات در لایه‌های مختلف شبکه، اتلاف اطّلاعات نداشته باشند. همچنین لازم است ورودی‌های متفاوت را در لایه‌های مختلف به‌صورت متمایزی بیان و به‌همین ترتیب در لایه‌های خروجی شبکه نیز بتوانند آنها را به‌صورت متمایزی بازسازی کنند. درصورتی‌که دو نمونهٔ 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 کند. لذا بخشی از اطّلاعات ورودی از دست رفته است.

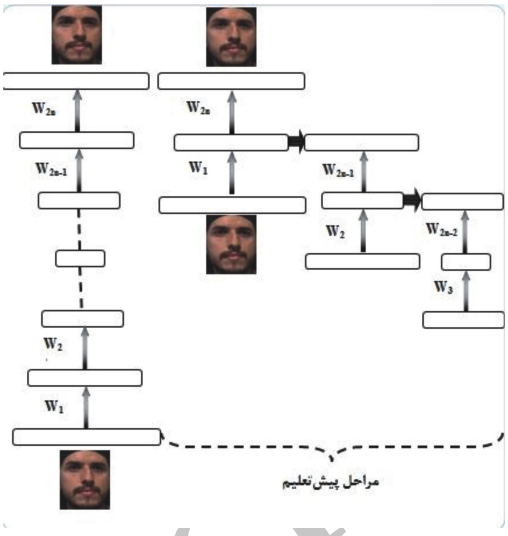
شبکه‌های عصبی عمیق که در نورون‌های آنها از تابع فعالیت پله‌ای استفاده شده باشد، با توجه به اینکه این نورون‌ها در فضای ورودی خود یک ابرصفحه ایجاد می‌کنند، نمونه‌های ورودی را در دو طبقه متمایز در خروجی نورون تفکیک می‌نمایند. چنین شبکه‌هایی برای اینکه قادر باشند تمامی نمونه‌های متمایز را در خروجی به‌صورت متمایز بیان کنند لازمه‌اش آن است که در همه لایه‌ها تمامی نمونه‌های ورودی با کدهای متمایزی بیان شده باشند. برای این‌منظور ضروریست که در فضای ورودی، بین هر دو نمونه 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 نمونه یک ناحیه مجزا توسط ابرصفحات آن لایه شکل گرفته باشد تا در هیچ لایه‌ای دو نمونهٔ مجزا با یک بیان واحد ارائه نشوند. بر اساس این تحلیل، برای مسئله تعلیم DBNN از روش شکستن مسئلهٔ پیچیدهٔ یادگیری شبکه‌های عصبی با ساختارهای عمیق به چند زیرمسئلهٔ یادگیری شبکه با ساختارهای ساده‌تر و حل مسئلهٔ همگرایی آنها و سپس استفاده از آنها به‌عنوان مقادیر اولیه برای DBNN استفاده می‌شود.

در این روش شبکهٔ عصبی با ساختار عمیق دارای $2n + 1$ لایهٔ نورون به n شبکهٔ عصبی با ۲ لایهٔ نورون شکسته می‌شود و هر یک از شبکه‌ها به‌گونه‌ای تعلیم داده می‌شوند که کمترین اتلاف اطّلاعات ورودی را داشته باشند و خروجی تا حدّ امکان به‌صورت متمایزی بازسازی شود. البته در اینجا به‌لحاظ استفاده از تعلیم شبکه بر پایه گرادیان و الگوریتم پس‌انتشار خطا، از تابع فعالیت نرم به‌جای تابع پله‌ای استفاده می‌شود که مشتق‌پذیر باشد. لیکن در اینجا نیز درعمل تمایزپذیری بهتر نمونه‌ها در لایه‌های مختلف و لایهٔ خروجی شبکهٔ عمیق نیز تا حدودی در گرو موقعیت ابرصفحات مربوط به توابع پله‌ای متناظر با این توابع نرم است. به این معنی که توابع غیرخطی نرم، مانند سیگموید درصورتی‌که در حین تعلیم به‌گونه‌ای تنظیم شوند که بین نمونه‌های متمایز قرار گیرند، قدرت تمایز بهتری را از خود ارائه می‌دهند؛ لذا هر یک از شبکه‌های کوچک‌تر بر مبنای کمینه‌سازی خطای جمعی و بیشینه‌سازی متمایزگری نمونه‌های ورودی خود به‌صورت جداگانه تعلیم می‌بینند. از این طریق وزن‌های نورون‌های لایه‌های مختلف شبکه از یک روش پیش‌تعلیم به‌نسبه کارایی بهره‌مند می‌شوند. این روش پیش‌تعلیم، وزن‌های اوّلیهٔ DBNN را به موقعیتی منتقل می‌کند که از بسیاری از کمینه‌های موضعی ممکن رهیده

[1] Restricted Boltzmann Machine

Archive of SID
www.SID.ir

باشد؛ لذا هم به کیفیت بازسازی بهتر ورودی در خروجی منجر می‌شود و هم در کاربرد استخراج مؤلّفه‌های غیرخطی، قادر می‌شود تا مؤلّفه‌هایی با کیفیت بهتر را استخراج کند.



(شکل ۴): نحوه تجزیه DBNN به BNNهای یک لایه پنهان

در شکل (۴) نحوهٔ شکسته‌شدن DBNN به تعدادی BNN، یک لایهٔ پنهان نشان داده شده است. برای طی مراحل پیش‌تعلیم، اوّلین BNN را در شکل (۴) در نظر بگیرید. وزن‌های بخش نگاشت و نگاشت معکوس این شبکه به‌ترتیب W_1 در اولین و W_{2n} در آخرین لایهٔ ساختار DBNN می‌باشند. این شبکه با دادگان تعلیم به‌صورت خودانجمنی تعلیم داده می‌شود. بدین‌ترتیب ماتریس وزن‌های اولیه لایه‌های اول (W_1) و آخر (W_{2n}) DBNN به‌دست می‌آید. سپس تصویر دادگان ورودی در فضای لایهٔ گلوگاه شبکه اوّل محاسبه و به‌عنوان ورودی BNN دوم برای تعیین مقادیر اوّلیهٔ وزن‌های W_2, W_{2n-1} مورد استفاده قرار می‌گیرد. این روند به‌همین ترتیب ادامه می‌یابد.

از آنجاکه مراحل پیش‌تعلیم روی این BNNهای یک‌لایهٔ پنهان صورت می‌گیرد، ابتدا لازم است که برای هر BNN یک تابع هزینه تعریف شود. این تابع به‌صورت تابع خطا رابطهٔ (۱) تعریف می‌شود.

$$E_i^L=\frac{1}{2}\sum_{k=1}^{D_i}\left(Y_{i-1}(k)-Z_i(k)\right)^2 \qquad (۱)$$

E_i^L در رابطهٔ (۱) خطای خودانجمنی BNN iام است که برابر با خطای بازسازی ورودی در لایه خروجی است. در این رابطه Y_{i-1} تصویر دادگان در لایهٔ گلوگاه BNN $(i-1)$ام مطابق رابطهٔ (۲) می‌باشد. به‌عبارتی تصویر دادگان ورودی در فضای لایهٔ گلوگاه BNN $(i-1)$ام به‌عنوان ورودی و خروجی مطلوب BNN iام به‌کار می‌رود. همچنین Y_0، بردار ورودی BNN اول، برابر با X همان بردار ورودی DBNN است. Z_i که خروجی BNN iام است در رابطهٔ (۳) تعریف شده و D_i نیز تعداد نورون‌های خروجی BNN iام است.

$$Y_{i-1}=f(W_{i-1}Y_{i-2}-b) \qquad i>1 \qquad (۲)$$

$$Z_i=f(W_{2n-i+1}Y_i-b) \qquad i>0 \qquad (۳)$$

در روابط فوق b مقدار سطح آستانه را برای تابع فعالیت نورون‌ها تعیین می‌کند. به‌منظور بهینه‌سازی توابع هزینه، از قانون دلتا استفاده می‌شود.

لذا اصلاح وزن‌های هر BNN مطابق رابطهٔ (۴) در جهت عکس گرادیان خطای تعریف شده باید صورت گیرد. در هر مرحله از پیش‌تعلیم مقادیر اولیه برای دو مجموعه وزن W_i و DBNN W_{2n-i+1} تعیین می‌شود.

$$\Delta W=-\eta\frac{\partial E}{\partial W} \qquad (۴)$$

لذا برای به‌روزرسانی W_{2n-i+1} خوهیم داشت:

$$\Delta W_{2n-i+1}=-\eta\frac{\partial E_i^L}{\partial W_{2n-i+1}}=\eta\delta_z Y_i \qquad (۵)$$

که δ_z در آن عبارتست از:

$$\delta_z=f'(Z_i)(Z_i-Y_{i-1}) \qquad (۶)$$

همچنین برای W_i نیز از رابطهٔ (۷) استفاده می‌شود که δ_y در رابطه (۸) تعریف شده است.

$$\Delta W_i=\eta\delta_y Y_{i-1} \qquad (۷)$$

$$\delta_y=f'(Y_i)\delta_z W_{2n-i+1} \qquad (۸)$$

باید در نظر داشت که این تعلیم لایه‌به‌لایه را می‌توان به هر میزان مورد نیاز تکرار کرد. می‌توان نشان داد که روش ارائه‌شده برای پیش‌تعلیم وزن‌ها، روش بسیار کارایی است که می‌تواند ارتباطات مهم بین داده‌های لایهٔ آشکار را به‌خوبی شناسایی کرده، ضبط نماید و امکان بازسازی مناسب داده‌های اولیه را نیز فراهم کند.

بعد از انجام مراحل پیش‌تعلیم، ادامهٔ فرآیند مشابه شبکهٔ عصبی خودانجمنی ساده انجام می‌شود. به این‌منظور وزن‌های به‌دست آمده به‌عنوان وزن‌های اولیه در DBNN

Archive of SID
www.SID.ir

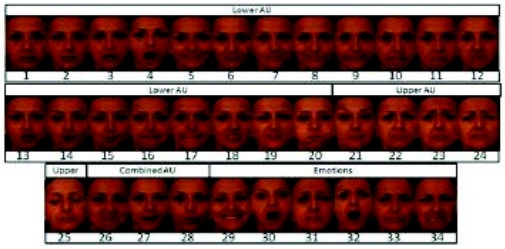
یک‌پارچه منظور شده و به کمک عملیات پس‌انتشار خطا مقادیر دقیق‌تر ماتریس‌های وزن شبکه به‌دست می‌آید.

۵. پیاده‌سازی و نتایج

جهت ارزیابی روش پیش‌تعلیم لایه‌به‌لایه، کارایی آن در تعلیم DBNN برای استخراج مؤلّفه‌های اساسی غیرخطی چهره مورد بررسی قرار گرفته است که در ادامه به شرح آن خواهیم پرداخت. ابتدا دادگان چهره مورد ارزیابی معرفی می‌شوند و سپس ساختار DBNN و نحوهٔ تعلیم آن شرح داده می‌شود. در ادامه آزمایش‌هایی در جهت نشان‌دادن قابلیت این روش پیش‌تعلیم صورت می‌گیرد. در این آزمایش‌ها به تأثیر آن در بهبود سرعت هم‌گرایی و قابلیت تعمیم‌دهی پرداخته می‌شود. همچنین مؤلّفه‌های استخراج شده توسط لایه گلوگاه DBNN ارزیابی می‌شوند. در این راستا کیفیت تصاویر بازسازی‌شده توسط آنها مقایسه می‌شود. همچنین به‌منظور بررسی قابلیت این روش در استخراج مؤلّفه‌های متمایزکننده از داده، با استفاده از مؤلّفه‌های استخراج‌شده در لایهٔ گلوگاه به بازشناسی افراد با استفاده از یک طبقه‌بند پرداخته می‌شود. درنهایت کارایی روش پیش‌تعلیم لایه‌به‌لایه در مقایسه با دو روش پیش‌تعلیم مشرف به هدف و تجزیه به ماشین‌های بولتزمان مورد بررسی قرار می‌گیرد.

۵-۱- پایگاه داده

در این تحقیق از دادگان بسفروس[1] (Sarvan et al., 2008) استفاده شده است. این پایگاه شامل تصاویر چهرهٔ سه‌بعدی و دوبعدی از ۱۰۵ سوژه است که در دانشگاه بگازیسی[2] ترکیه جمع‌آوری شده است. دادگان شامل مجموعه‌ای غنی از حالات چهره، چرخش سر و انواع مختلف انسداد است که مجموعه مربوط حالات مختلف چهره برای ارزیابی در این مقاله مورد استفاده قرار گرفته است. شکل (۵) این حالت‌ها را برای فرد ۴۳ام نشان می‌دهد.



(شکل ۵): تصاویر مربوط به حالت‌های مختلف چهره برای فرد ۴۳ام از پایگاه داده بسفروس (Sarvan et al., 2008).

تصاویر ۹۵ نفر از این مجموعه به‌طور تصادفی جهت بررسی انتخاب شد. این تصاویر به‌صورت سیاه و سفید با ۲۵۶ سطح خاکستری مورد استفاده قرار گرفت. همچنین جهت کاهش بار محاسبات وضوح این تصاویر به ۱۱۴×۹۲ کاهش داده شد. از آنجا که تعداد تصاویر مربوط به حالات مختلف چهره برای تمام افراد یکسان نیست، دوازده حالت مشترک بین تمام افراد برای آموزش و بقیه حالت‌های چهره برای آزمون کنار گذاشته شد. درنهایت مجموعه‌های آموزش و آزمون به‌ترتیب شامل ۱۱۲۶ و ۱۴۴۳ تصویر شد.

۵-۲- پیش‌پردازش دادگان چهره

هر تصویر چهره به‌صورت یک آرایهٔ دوبعدی M×N از مقادیر روشنایی در نظر گرفته می‌شود؛ اما از آنجاکه در ورودی شبکه، این آرایه باید به‌صورت یک‌بعدی ارائه شود، هر تصویر به یک بردار MN×1 تبدیل می‌شود. همچنین مقدار هر پیکسل تصویر بر بالاترین سطح روشنایی (عدد ۲۵۵) تقسیم می شود تا مقادیر پیکسل‌های هر تصویر چهره به محدوده صفر تا یک به‌هنجار شوند.

۵-۳- تعلیم DBNN

ساختار DBNN جهت استخراج مؤلّفه‌های غیرخطی تصاویر چهره در جدول (۱) معرفی شده است.

(جدول ۱): پارامترهای DBNN

تعداد لایه‌های پنهان	۵
تعداد نورون لایه پنهان	۴۰۰-۲۰۰-۱۰۷-۲۰۰-۴۰۰
تابع نورون‌های لایه پنهان	غیرخطی تک قطبی
تابع نورون‌های لایه خروجی	خطی
ضریب یادگیری	۰/۰۰۱
ضریب گشتاور	۰/۷

برای تعلیم این شبکه دو روش استفاده شده است. در روش اول مقادیر اولیه وزن‌ها به‌صورت تصادفی قرار داده شد. درحالی‌که در روش دوم وزن‌ها با استفاده از روش پیش‌تعلیم

[1] Bosphorus
[2] Bogazici

Archive of SID
www.SID.ir

لایه‌به‌لایهٔ پیشنهادی مقداردهی اولیه شد. بدین‌منظور DBNN هفت لایه به سه BNN تک لایه شکسته شد. با توجه به اینکه دادگان حاوی تنوعات مختلف چهره افراد است، به‌منظور سرعت بخشیدن به تعلیم با توجه به ایدهٔ مطرح شده در (Bengio et al., 2009) برای تعلیم این BNNها فقط از تصاویر طبیعی افراد استفاده شد و تنوعات آنها در مرحله تنظیم دقیق وزن‌ها آموزش داده شد. این روش کمک می‌کند در مرحله پیش‌تعلیم، شبکه مؤلّفه‌های ساده‌تر (مؤلّفه‌های متمایزکننده چهره افراد) و مؤلّفه‌های سطح بالاتر مانند چهره فرد با لبخند را در مرحلهٔ بعد بیاموزد.

ابتدا BNN اوّل با ساختار (۱۰۴۸۸-۴۰۰-۱۰۴۸۸) تعلیم داده شد و سپس مؤلّفه‌های استخراج شده در لایهٔ گلوگاه آن با بعد ۴۰۰ برای تعلیم BNN دوم با ساختار (۴۰۰-۲۰۰-۴۰۰) به‌کار رفت و برای BNN سوم هم به‌همین ترتیب عمل شد. سپس وزن‌های حاصل از مرحله پیش‌تعلیم در ساختار اصلی DBNN جایگزین شد و تصاویر طبیعی و تنوّعات آنها برای تنظیم دقیق وزن‌ها تعلیم داده شد.

۵-۴- تأثیرات روش پیش‌تعلیم لایه‌به‌لایه در تعلیم DBNN

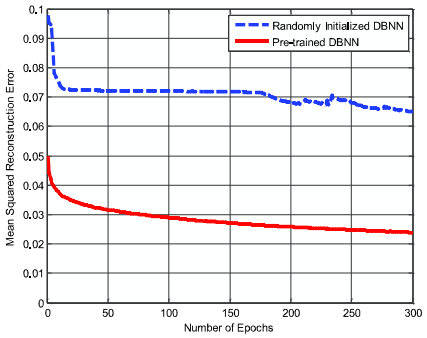
در ادامه به مقایسهٔ نتایج حاصل از تعلیم DBNN با دو روش مقداردهی اوّلیه برای وزن‌ها، به‌صورت تصادفی و پیش‌تعلیم لایه‌به‌لایه، در تعداد تکرار یکسان تعلیم پرداخته می‌شود. جهت ارزیابی بهتر، وزن‌های DBNN پنج مرتبه به‌صورت تصادفی مقداردهی شد و تعلیم صورت گرفت. لذا میانگین نتایج این پنج شبکه برای مقداردهی تصادفی گزارش می‌شود.

۵-۴-۱- همگرایی سریع‌تر

در این بخش به‌منظور بررسی کارایی روش پیش‌تعلیم لایه‌به‌لایه در همگرا کردن تعلیم DBNN و سرعت بخشیدن به آن، مجذور میانگین خطای بازسازی[1] تصاویر تعلیم در طی آموزش شبکه ارزیابی شده است. این خطا با رابطهٔ (۹) تعریف می‌شود.

$$E_{MSRE}=\sqrt{\frac{\sum_{j=1}^{n}\sum_{i=1}^{p}\left(I_{ij}-Z_{ij}\right)^2}{pn}} \qquad (۹)$$

در رابطهٔ (۹)، I_{ij} مقدار واقعی هر پیکسل و Z_{ij} خروجی DBNN به‌ازای آن پیکسل است. p و n نیز به‌ترتیب تعداد پیکسل هر تصویر و تعداد تصاویر تعلیم هستند. در نمودار شکل (۶) میانگین این خطا برای پنج‌بار مقداردهی تصادفی و همچنین مقداردهی با روش پیش‌تعلیم لایه‌به‌لایه در تعلیم DBNN آورده شده است. همان‌طور که نمودار شکل (۶) نشان می‌دهد، در تعلیم DBNN برای تعداد تکرار یکسان، پیش‌تعلیم وزن‌ها خطای آموزش پایین‌تری در مقایسه با مقداردهی اولیه تصادفی نتیجه می‌دهد. لذا روش پیش‌تعلیم لایه‌به‌لایه به همگرایی سریع‌تر تعلیم به میزان قابل توجهی کمک می‌کند. قابل ذکر است که در این بخش در آزمایش‌های مکرر با مقداردهی تصادفی وزن‌ها، تعلیم همگرا نشد و نتایج گزارش‌شده مربوط به موارد همگراشده است.



(شکل ۶): نمودار تغییرات میانگین خطای تعلیم DBNN برای ۵ بار مقداردهی تصادفی (—) و مقداردهی با روش پیش‌تعلیم لایه‌به‌لایه (—)

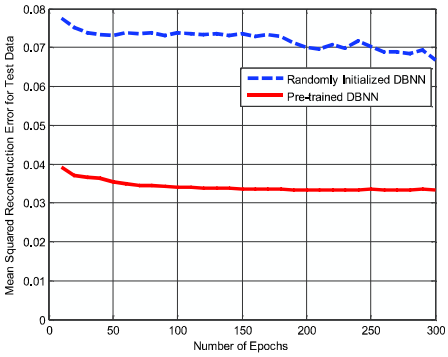
۵-۴-۲- تعمیم بهتر

برای ارزیابی قابلیت تعمیم DBNN، لازم است کارایی شبکه DBNN با مقداردهی اولیه توسط روش پیش‌تعلیم لایه‌به‌لایه در استخراج مؤلّفه‌های غیرخطّی تصاویر آزمون بررسی شود. بدین‌منظور خطای بازسازی تصاویر آزمون به‌صورت کمّی و همچنین تصاویر بازسازی‌شده به‌طور کیفی مقایسه می‌شوند.

در نمودار شکل (۷) در طی تعلیم DBNN، E_{MSRE} برای دادگان آزمون به‌صورت برخط، برای پنج‌بار مقداردهی اولیه تصادفی و مقداردهی با روش پیش‌تعلیم لایه‌به‌لایه آورده شده است. همان‌طور که مشاهده می‌شود E_{MSRE} برای دادگان آزمون با روش پیش‌تعلیم لایه‌به‌لایه به‌طور محسوسی پایین‌تر است. این نتیجه بیان‌گر این است که اعمال روش پیش‌تعلیم لایه‌به‌لایه منجر به تعمیم‌پذیری بهتر DBNN می‌شود.

[1] Mean Squared Reconstruction Error

Archive of SID
www.SID.ir



(شکل ۷): نمودار تغییرات میانگین مجذور خطای بازسازی هر پیکسل برای دادگان آزمون در طی تعلیم DBNN برای ۵ بار مقداردهی تصادفی (--) و مقداردهی با روش پیش‌تعلیم لایه‌به‌لایه (—)

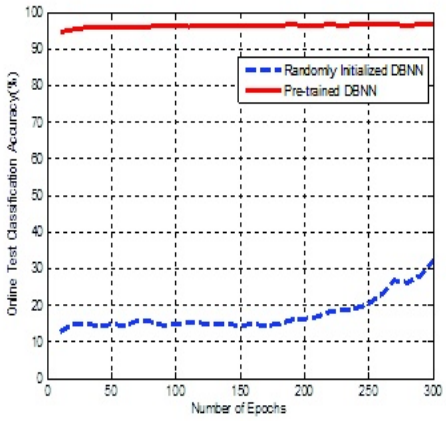
برای مقایسه کیفی روش پیشنهادی با روش تصادفی در استخراج مؤلّفه‌های بهینه‌تر توسط DBNN، تصاویر شش فرد از مجموعه آزمون به‌صورت تصادفی انتخاب شده است که سطر اول در شکل (۸) این تصاویر را نشان می‌دهد.

(شکل ۸): از بالا به پایین: تصاویر تصادفی از مجموعه آزمون، تصاویر بازسازی شده از ۱۰۷ مؤلّفه استخراج شده توسط DBNN با مقادیر اولیه تصادفی برای وزن‌های شبکه، تصاویر بازسازی شده از ۱۰۷ مؤلّفه استخراج شده توسط DBNN با اعمال روش پیش‌تعلیم لایه‌به‌لایه برای وزن‌های شبکه

همان‌طور که در (جدول ۱) آورده شده است DBNN تعریف‌شده در لایهٔ گلوگاه دارای ۱۰۷ نورون است که در مقایسه با بعد تصاویر (۱۱۴×۹۲) حدود ۹۸ برابر کمتر است. لذا DBNN هرچه مؤلّفه‌های بهینه‌تری را استخراج کند کیفیت تصاویر بازسازی‌شده بهتر خواهد بود. سطر دوم در شکل (۸) تصاویر بازسازی‌شده را با یکی از شبکه‌های مقداردهی‌شده به‌صورت تصادفی نشان می‌دهد. در سطر سوم تصاویری که آورده شده است مربوط به خروجی شبکهٔ مقداردهی‌شده با روش پیش‌تعلیم لایه‌به‌لایه است. در مقایسهٔ تصاویر سطر دوم و سوم مشاهده می‌شود که با وجود تعداد تکرار تعلیم یکسان، تصاویر بازسازی‌شده توسط DBNN با پیش‌تعلیم از وضوح بهتری برخوردار هستند.

۵-۴-۳- استخراج مؤلّفه‌های متمایزکننده

در این بخش به بازشناسی افراد با استفاده از مؤلّفه‌های استخراج‌شده توسط DBNN پرداخته می‌شود. برای بازشناسی از طبقه‌بند KNN[1] استفاده شده است. مؤلّفه‌های استخراج‌شده برای تصاویر آموزش و آزمون به‌ترتیب دادگان آموزش و آزمون این طبقه‌بند می‌باشند. نمودار شکل (۹) تغییرات درصد صحت بازشناسی افراد را در طی تعلیم DBNN را برای دادگان آزمون نشان می‌دهد. با توجه به این نمودار، روش پیش‌تعلیم لایه‌به‌لایه در راستای استخراج مؤلّفه‌های متمایزکننده به DBNN کمک می‌کند. به‌طوری‌که درصد صحت بازشناسی افراد در DBNN با پیش‌تعلیم در مقایسه با مقداردهی تصادفی تفاوت معناداری دارد.



(شکل ۹): نمودار درصد صحت بازشناسی مؤلّفه‌های تصاویر آزمون توسط KNN در طی تعلیم DBNN برای پنج بار مقداردهی تصادفی (--) و مقداردهی با روش پیش‌تعلیم لایه‌به‌لایه (—)

با دقّت در نرخ بازشناسی افراد برای DBNN با پیش‌تعلیم در نمودار شکل (۹) مشاهده می‌شود که با جلورفتن مسیر تعلیم، درصد صحّت بازشناسی تغییر زیادی نمی‌کند. این مسأله نشان می‌دهد که در مراحل پیش‌تعلیم این وزن‌ها در محدوده مناسب قرار گرفته‌اند.

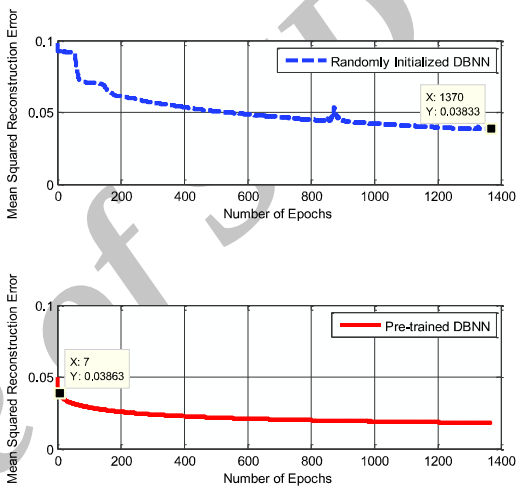
[1] K-Nearest Neighbor

Archive of SID
www.SID.ir

۵-۵- تعمیم بهتر با خطای تعلیم یکسان

با توجّه به اینکه در طی تعداد تکرارهای یکسان تعلیم، میزان خطای تعلیم برای مقداردهی اولیه تصادفی و با پیش‌تعلیم یکسان نبود در این بخش تعلیم ادامه داده شد تا زمانی‌که خطا برای یکی از DBNNها با مقداردهی اولیه تصادفی وزن‌ها با خطای تعلیم DBNN با پیش‌تعلیم در یکی از تکرارها برابر گردد.

همان‌طور که نمودار شکل (۱۰) نشان می‌دهد در تکرار ۱۳۷۰، DBNN با مقداردهی اولیه تصادفی به خطای ۰/۰۳۸ معادل خطای تکرار ۷ام DBNN با پیش‌تعلیم رسیده است. مجموعه وزن‌های این دو تکرار جهت ارزیابی در این بخش مورد استفاده قرار گرفته است.



(شکل ۱۰): نمودار تغییرات خطای تعلیم DBNN برای مقداردهی تصادفی (--) و مقداردهی با روش پیش‌تعلیم لایه‌به‌لایه (—)

در جدول (۲) خطای بازسازی تصاویر آزمون مطابق رابطهٔ (۹) برای DBNN با هر دو روش تعلیم محاسبه شده است. در جدول (۳) نیز خطای بازشناسی تصاویر آزمون آورده شده است. برای محاسبهٔ خطای بازشناسی مانند بخش (۵-۴-۳) مؤلّفه‌های استخراج شده توسط DBNN به طبقه‌بند KNN داده شده است و برچسب افراد تعیین می‌شود.

(جدول ۲): مقایسه میانگین مجذور خطای بازسازی هر پیکسل (E_{MSRE}) در تصاویر آزمون برای DBNN با دو روش مقداردهی اولیه تصادفی و پیش‌تعلیم لایه‌به‌لایه با خطای تعلیم یکسان

DBNN با مقادیر اولیه تصادفی برای وزن‌ها	۰/۰۴۵۳
DBNN با پیش‌تعلیم لایه‌به‌لایه	۰/۰۳۹۱

(جدول ۳): مقایسه خطای بازشناسی افراد در تصاویر آزمون برای DBNN با دو روش مقداردهی اولیه تصادفی و پیش‌تعلیم لایه‌به‌لایه با خطای تعلیم یکسان

DBNN با مقادیر اولیه تصادفی برای وزن‌ها	۸۴/۰۶٪
DBNN با پیش‌تعلیم لایه‌به‌لایه	۹۴/۷۲٪

همان‌طور که مشاهده می‌شود با وجود خطای تعلیم یکسان با به کارگیری روش پیش‌تعلیم لایه‌به‌لایه برای مقداردهی اولیه وزن‌های DBNN، خطای بازسازی هر پیکسل ۰/۰۰۶ معادل ٪۱۳/۶۹ کاهش و درصد صحّت بازشناسی افراد حدود ٪۱۰ رشد داشته است.

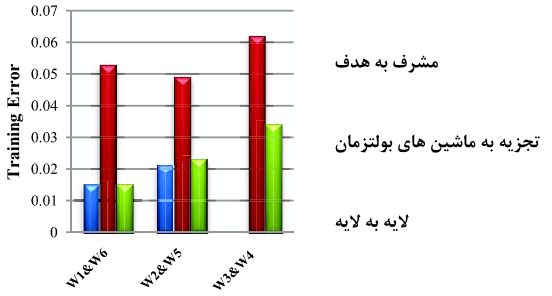
۵-۶- مقایسه روش پیش‌تعلیم لایه‌به‌لایه با برخی روش‌های پیش‌تعلیم

در این آزمایش کارایی روش پیش‌تعلیم لایه‌به‌لایه در مقایسه با دو روش پیش‌تعلیم مشرف به هدف (نژادقلی،۱۳۸۳،۱۳۹۱) و تجزیه به ماشین‌های بولتزمان (Hinton & Salakhutdinov, 2006) مورد بررسی قرار می‌گیرد. تصاویر آموزش در این آزمایش همان بردارهای ۱۰۴۸۸تایی پیکسل‌های تصاویر چهره مجموعه تعلیم هستند که در (بخش ۵-۱) معرفی شدند. ساختار DBNN هم مطابق (جدول ۱) است. برای پیش‌تعلیم با روش لایه‌به‌لایه و تجزیه به ماشین‌های بولتزمان، ساختار DBNN به‌ترتیب به سه BNN یک لایهٔ پنهان و سه RBM تجزیه می‌شود و پیش‌تعلیم در سه مرحله صورت می‌گیرد. در روش تجزیه به ماشین‌های بولتزمان واحدهای قابل رؤیت باینری با توجه به ماهیت پیوستهٔ مقادیر ورودی برای اولین ماشین بولتزمان، با واحدهای گوسی (Hinton & Salakhutdinov, 2006) جایگزین شده است؛ ولی نورون‌های پنهان آن همچنان باینری است. برای روش مشرف به هدف، پیش‌تعلیم در دو مرحله با دو BNN به‌ترتیب با یک و دو لایه پنهان انجام می‌شود. در شکل (۱۱) خطای تعلیم در طی مراحل پیش‌تعلیم بعد از تعداد تکرارهای مساوی روش‌های پیش‌تعلیم آورده شده است.

در مرحلهٔ اوّل که وزن‌های لایه‌های اول و آخر مقداردهی اوّلیه می‌شوند، روش مشرف به هدف و لایه‌به‌لایه یکسان

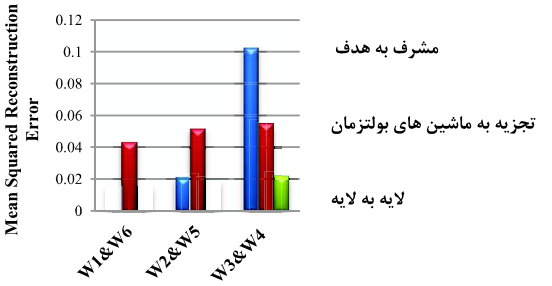
Archive of SID
www.SID.ir

عمل می‌کنند. امّا در مورد روش تجزیه به ماشین‌های بولتزمان مشاهده می‌شود که خطای تعلیم آن نسبت به دو روش دیگر بالاتر است. در مراحل بعد پیش‌تعلیم نیز به‌طور



(شکل ۱۱): مقایسه خطای تعلیم تصاویر آموزش در طی مراحل پیش‌تعلیم وزن‌های DBNN با ۵۰۰ بار تکرار در هر مرحله

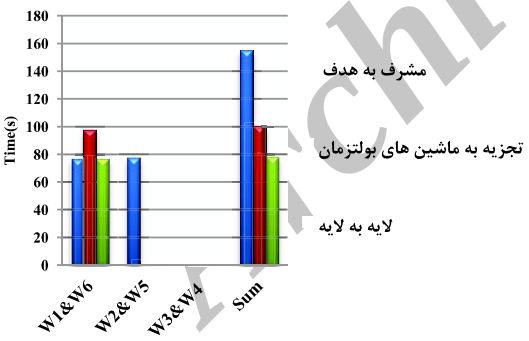
تقریبی نتایج همین روند را دنبال می‌کنند. با توجّه به این بررسی دو روش لایه‌به‌لایه و مشرف به هدف در مقایسه با روش تجزیه به ماشین‌های بولتزمان در طی مراحل پیش‌تعلیم، با تعداد تکرار مساوی به خطای تعلیم پایین‌تری دست می‌یابند. همچنین روش لایه‌به‌لایه در مقایسه با مشرف به هدف با توجه به اینکه وزن‌های لایه‌های سوم و چهارم را نیز مقداردهی اولیه می‌کند، نقطه شروع مناسب‌تری برای وزن‌های شبکه عمیق فراهم می‌آورد.



(شکل ۱۲): مقایسه خطای بازسازی تصاویر آموزش در طی مراحل پیش‌تعلیم وزن‌های DBNN با ۵۰۰ بار تکرار در هر مرحله

در شکل (۱۲) E_{MSRE} برای تصاویر آموزش در طی مراحل پیش‌تعلیم بعد از تعداد تکرارهای مساوی روش‌های پیش‌تعلیم آورده شده است. همان‌طور که در این نمودار مشاهده می‌شود خطای بازسازی تصاویر مانند خطای آموزش برای دو روش مشرف به هدف و لایه‌به‌لایه در طی مراحل پیش‌تعلیم پایین‌تر است. مقایسه مقادیر حاصل از دو روش مشرف به هدف و لایه‌به‌لایه در مرحله آخر پیش‌تعلیم نشان می‌دهد که پیش‌تعلیم مجزای لایه‌ها به نتیجهٔ بهتری منجر می‌شود.

در نمودارهای شکل (۱۳) متوسط مدّت‌زمان اجرای مراحل پیش‌تعلیم وزن‌های DBNN برای هر بار تکرار آموزش آورده شده است. در طی مراحل پیش‌تعلیم برای روش‌های تجزیه به ماشین‌های بولتزمان و لایه‌به‌لایه با توجه به اینکه با حرکت به سمت لایه‌های عمیق‌تر تعداد نورون‌های لایه‌ها نیز کاهش می‌یابد، پیچیدگی RBMها و BNNها نیز کمتر می‌شود. لذا همان‌طور که مشاهده می‌شود، زمان تعلیم نیز کاهش می‌یابد. امّا در هر سه مرحله این زمان برای روش پیش‌تعلیم لایه‌به‌لایه در مقایسه تجزیه به ماشین‌های بولتزمان کمتر است. بر خلاف این دو روش برای روش مشرف به هدف از آنجا که با جلورفتن مراحل پیش‌تعلیم تعداد لایه‌های شبکه مورد آموزش نیز افزایش می‌یابد و در عین حال تعداد نورون‌های خروجی و ورودی همچنان ثابت و برابر با بعد دادهٔ ورودی است، لذا این زمان افزایش می‌یابد. به‌طوری‌که متوسط زمان تمام مراحل پیش‌تعلیم برای روش مشرف به هدف، حدوداً دو برابر روش لایه‌به‌لایه است.



(شکل ۱۳): مقایسه متوسط مدت زمان مراحل پیش‌تعلیم وزن‌های DBNN برای هر بار تکرار آموزش (ثانیه)

۶. بحث و نتیجه‌گیری

در برخی موارد شبکه یک لایه پنهان در استخراج مؤلّفه‌های دادهٔ ورودی موفّق عمل نمی‌کند؛ چراکه برای داده‌های

Archive of SID
www.SID.ir

پیچیده‌تر با بعد بالا مانند تصاویر چهره تبدیل ناگهانی ویژگی‌های دادهٔ ورودی به تعداد مؤلفّهٔ محدود در لایهٔ گلوگاه، کیفیت بازسازی را بسیار مخدوش کرده و خطای شبکه را به میزان قابل توجهی افزایش می‌دهد. همچنین پردازش‌های غیرخطی پیچیدهٔ عمیق با یک لایهٔ قابل پیاده‌سازی نیستند . لذا نیاز به ساختارهای عمیق‌تر شبکه‌های عصبی به‌منظور استخراج مؤلّفه برای داده‌های پیچیده‌تر است تا امکان پردازش‌های غیرخطّی چندمرحله‌ای متوالی فراهم شود. این درحالیست که با افزایش تعداد لایه‌های شبکه‌های عصبی، همگرایی تعلیم آن به‌دلیل افزایش کمینه‌های موضعی با مشکل مواجه می‌شود. به‌منظور حلّ این مشکل در این مقاله روش پیش‌تعلیم لایه‌به‌لایه جهت مقداردهی مناسب وزن‌های شبکه ارائه شد. این روش در تعلیم BNN اوّل مشابه روش مشرف به هدف (نژادقلی، ۱۳۸۳) عمل می‌کند. امّا برخلاف روش مشرف به هدف که به رشددادن ساختار BNN اوّل می‌پردازد، خروجی لایه گلوگاه آن‌را برای تعلیم BNN دوم به‌کار می‌برد که چند مزیت دارد. اوّل اینکه در هر مرحله یک BNN یک لایهٔ پنهان باید تعلیم داده شود؛ لذا سرعت پیش‌تعلیم در مقایسه با مشرف به هدف که نیاز به تعلیم ساختار رشد داده شده را دارد، به‌خصوص در لایه‌های عمیق‌تر، بیشتر است که آزمایش‌های انجام شده تأییدکنندهٔ این مسئله هستند. همچنین با توجه به روند کاهشی تعداد نورون‌های DBNN، استفاده از خروجی لایهٔ گلوگاه BNN قبلی برای تعلیم BNN جدید باعث می‌شود بعد دادهٔ تعلیم BNN کاهش یابد لذا سرعت افزایش می‌یابد، درحالی‌که در مراحل رشد ساختار در روش مشرف به هدف از داده اصلی استفاده می‌شود که دارای بعد بالایی است. علاوه‌بر این با رشد ساختار، امکان گیر افتادن در کمینه‌های موضعی برای روش مشرف به هدف به دلیل پیچیده‌ترشدن شبکه، افزایش می‌یابد؛ لذا همگرایی کاهش می‌یابد.

روش پیش‌تعلیم لایه‌به‌لایه در مقایسه با روش تجزیه شبکه به ماشین‌های بولتزمان (Hinton & Salakhutdinov, 2006)، به‌طور تقریبی مشابه عمل می‌کند با این تفاوت که در این روش به‌جای تجزیه به ماشین‌های بولتزمان، DBNN به تعدادی BNN متناظر تجزیه می‌شود. استفاده از BNN در مقایسه با ماشین بولتزمان برتری‌هایی دارد. اوّل این‌که تعلیم آن مشابه DBNN و با الگوریتم پس‌انتشار خطا صورت می‌گیرد. دوم این‌که تعلیم آن سریع‌تر است. به‌گونه‌ای‌که بررسی‌ها نشان داد که برای تعداد تکرار یکسان تعلیم، روش لایه‌به‌لایه به خطای تعلیم پایین‌تر در زمان کمتری دست می‌یابد. سوم این‌که به داده آموزشی کمتری نیاز دارد (نژادقلی، ۱۳۹۱).

کارایی روش پیش‌تعلیم لایه‌به‌لایه در همگرا کردن DBNN به‌منظور استخراج مؤلّفه‌های چهره مورد ارزیابی قرار گرفت. نتایج نشان داد با به‌کارگیری روش پیش‌تعلیم لایه‌به‌لایه نه‌تنها سرعت همگرایی تعلیم در حد چشم‌گیری افزایش می‌یابد، بلکه قابلیت تعمیم‌دهی شبکه در استخراج مؤلّفه‌های بهینه برای تصاویر آزمون جهت بازسازی و طبقه‌بندی نیز بهبود می‌یابد.

مراجع

سیّدصالحی، سیّدعلی، توحیدخواه، فرزاد، نژادقلی، ایثار، «افزایش کارایی بازشناخت الگوی شبکه‌های عصبی جلوسو از طریق توسعهٔ روش‌هایی برای دوسویه کردن عمل‌کرد آنها،» گزارش طرح مستقل پژوهشی، دانشگاه صنعتی امیرکبیر، ۱۳۸۳.

قاسمی، مهدیه، «استخراج مؤلّفه‌های مستقل غیرخطّی سیگنال گفتار،» پایان‌نامه کارشناسی ارشد بیوالکتریک، دانشکدهٔ مهندسی پزشکی، دانشگاه صنعتی امیرکبیر، ۱۳۸۴.

نژادقلی، ایثار، «آزمایش‌هایی جهت دوسویه کردن شبکه‌های عصبی،» گزارش تحقیقاتی پژوهشکدهٔ پردازش هوشمند علائم، بخش پردازش گفتار، ۱۳۸۳.

نژادقلی، ایثار، « مدل‌سازی نحوهٔ استخراج و بازترکیب ویژگی‌های ادراکی در مغز با لحاظ کردن تعاملات بین آنها،» پایان‌نامه دکترا بیوالکتریک، دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر، ۱۳۹۱.

Bengio, Y., “Evolving Culture vs Local Minima,” University de Montréal, technical report, 2012.

Archive of SID
www.SID.ir

Savran, A., Alyüz, N., Dibeklioğlu, H., Çeliktutan, O., Gökberk, B., Sankur, B., Akarun, L., "Bosphorus Database for 3D Face Analysis," BIOID 2008, Roskilde University, Denmark, May 2008.

Seow, M.J., "Learning as a nonlinear line of attraction for pattern association, classification and recognition," PhD. Thesis, Old Dominion University, 2006.

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.

سـیّده زهـره سیّدصـالحی مـدرک کارشناسی خـود را دررشـتهٔ مهندسـی پزشکی- بیوالکتریک از دانشگاه صنعتی امیرکبیـر در سـال ۱۳۸۳و کارشناسـی ارشد را در همان رشته از دانشکدهٔ فنی دانشگاه شاهد در سـال ۱۳۸۶ دریافـت نمـوده اسـت. وی از سال ۱۳۸۷ تا کنون دانشجوی دکترای دانشکدهٔ مهندسـی پزشکی دانشگاه صنعتی امیرکبیر است. زمینه‌های پژوهشی مورد علاقهٔ ایشان پردازش سیگنال با بهره‌گیری از روش‌های هوش مصنوعی و شبکه‌های عصبی مصنوعی است.

نشانی رایانامهٔ ایشان عبارت است از:

z.seyyedsalehi@aut.ac.ir

سیّدعلـی سیّدصـالحی مـدرک کارشناسی خود را در رشتهٔ مهندسی برق از دانشـگاه صـنعتی شـریف در سال ۱۳۶۱، کارشناسـی ارشـد را در همـان رشـته از دانشـگاه صـنعتی امیرکبیر در سال ۱۳۶۷ و دکترای خود را در رشتهٔ مهندسی برق- بیوالکتریک از دانشگاه تربیـت مـدرس در سـال ۱۳۷۴ دریافت کرده‌اند. ایشـان در حـال حاضـر دانشـیار دانشـکدهٔ

Bengio, Y., "Learning deep architectures for AI" in: Foundations and Trends in Machine Learning, pp. 1-127, 2009.

Bengio, Y., Louradour, J., Collobert, R. and Weston, J., "Curriculum Learning," in: Presented in International Conference on Machine Learning, 2009.

Daszykowski, B., Walczak, B., Massart, D.L., "A journey into low-dimensional spaces with autoassociative neural networks," Talanta, vol. 59, pp. 1059-1105, 2003.

Erhan, D., Manzagol, P.A., Bengio, Y., Bengio, S., and Vincent, P., "The difficulty of training deep architectures and the effect of unsupervised pre-training," AISTATS'09, pp. 153–160, 2009.

Erhan, D. , Bengio, Y. , Courville, A. , Manzagol, P.A. , Vincent, P. and Bengio, S. , "Why does unsupervised pre-training help deep learning?" Journal of Machine Learning Research, vol. 11, pp. 625-660, 2010.

Erhan, D. , "Understanding deep architectures and the effect of unsupervised pre-training," PhD Thesis,Universit´e de Montr´eal, (2011)

Hinton, G.E., and Salakhutdinov, R., "Reducing the dimensionality of data with neural networks," Science, vol. 313(5786), pp. 504–507, 2006.

Kramer, M.A., "Nonlinear principal component analysis using auto associative neural networks," AIChE J. 37, 223_243, 1991.

Krueger, K.A., and Dayan, P., "Flexible shaping: how learning in small steps helps," Cognition, vol. 110, pp. 380–394, 2009.

Larochelle, H., Erhan, D., Courville, A., Bergstra, J., and Bengio, Y., "An empirical evaluation of deep architectures on problems with many factors of variation," Proceedings of the 24th international conference on Machine learning, pp. 473_480, 2007.

Maaten, L.J.P., Postma, E.O., and Herik, H.J., "Dimensionality Reduction: A Comparative Review," Tilburg University Technical Report, TiCC-TR 2009-005, 2009.

Plath, N., "Extracting low-dimensional features by means of Deep Network Architectures," PhD. Thesis, Technische Universität Berlin, April 2008.

Archive of SID

www.SID.ir

مهندسی پزشکی دانشگاه صنعتی امیرکبیر هستند. زمینه‌های پژوهشی مورد علاقهٔ ایشان پردازش و بازشناسی گفتار، شبکه‌های عصبی مصنوعی و زیستی، مدل‌سازی عمل‌کرد مغز و پردازش خطی و غیرخطّی سیگنال است.
نشانی رایانامهٔ ایشان عبارت است از:
ssalehi@aut.ac.ir

Archive of SID

www.SID.ir