بهبود قدرت تعمیم شبکه های عصبی با تعیین تطبیقی ضرایب انتظام

سیّدمهدی فخرایی
استادیار

کارولوکس
استاد

بهرام سجادی بی ریا
دانشجوی کارشناسی ارشد

دانشکده مهندسی برق و کامپیوتر، دانشگاه تهران

چکیده

از مسائل اساسی در آموزش شبکه های عصبی مصنوعی حصول اطمینان از انجام تعمیم مناسب برای آن دسته از نمونه هایی که به هنگام آموزش به شبکه اعمال نشده اند، می باشد. وجود تعداد نمونه های آموزشی به گونه ای که به طور قابل ملاحظه ای بیش از تعداد وزن های غیروابسته شبکه باشد، می تواند تضمینی برای حصول قدرت تعمیم مناسب به شمار آید. در بسیاری از مسائل عملی، فقط مقدار کمی از داده های برچسب خورده جهت آموزش موجود می باشد، که موجب بروز مشکل برای راه حل هایی می گردد که معماری مناسب خود را بر اساس آنالیزهای مفصل و پیچیده نظری انتخاب نکرده اند. در نتیجه، روش هایی که توانایی تعلیم شبکه های بزرگ، با تعداد نسبتاً کم داده های آموزشی را به گونه ای داشته باشند که شبکه حاصل از آموزش دارای قابلیت تعمیم مناسب و کارآیی بالایی باشد، مورد توجه خاص قرار می گیرند. یکی از این روش ها کنترل پیچیدگی شبکه در طول فرآیند آموزش توسط استفاده از عبارت انتظام می باشد. در این مقاله ضمن استفاده از ضرایب انتظام متعدد، روشی جهت تغییر مقادیر این ضرایب در حین آموزش، صرفاً با استفاده از داده های آموزشی ارائه گردیده است.

Improvement of Neural Networks Generalization by Adaptive Regularization Coefficients

S. M. Fakhraie
Assistant Professor

C. Lucas
Professor

B. Sadjadi Biria
Student

Elec. Engineering Department, University of Tehran

Abstract

One of the main problems in the neural-network learning is obtaining a confidence level of generalizing to the samples which have never been seen by the network during the training process.

Existence of many more learning samples than the independent network weights, improves the generalization performance. However in many application areas there are only a few labeled training samples, which causes some difficulties for the solution and its generalization capability.

Thus, providing some solutions with the ability to train large networks by using a few labeled learning samples, to generalize well, is of great importance. One of these possible solutions is regularization, which controls the complexity of the network during training.

In this paper, we introduce an analytical solution which sets the multiregularization coefficient values by using only the learning samples during training.

کلمات کلیدی

شبکه عصبی، قدرت تعمیم، انتظام، ضرایب انتظام متعدد، ضرایب انتظام تطبیقی.

مقدمه

شبکه های عصبی مصنوعی ابزاری مناسب در کاربردهای شناسایی الگو و تقریب توابع بوده و با توسعه و گسترش شبکه به میزان کافی، تقریباً هر تابعی با دقت دلخواه قابل تقریب می باشد [۱] [۲] [۳]. مهمترین و مورد توجه ترین ویژگی شبکه های عصبی توانایی یادگیری از روی مثال است، بدین ترتیب که انتظار استخراج روند صحیح تولید کننده داده های آموزشی را از شبکه داریم، بدون اینکه مدل ریاضی خاصی از آن در دست باشد. به عبارت دیگر صرفاً توسط نمونه های موجود از رفتار مقصود اطلاعات لازم به شبکه انتقال می یابد. آنچه موجب نگرانی در یادگیری شبکه های عصبی می شود بروز پدیده برازش اضافی[۱] بر روی نوفه[۲] موجود در داده های آموزشی می باشد. در این صورت به جای آنکه روند حاکم بر تولید داده ها توسط شبکه کشف گردد، جزئیات موجود در داده های آموزشی فرا گرفته می شود و در نتیجه از قدرت تعمیم شبکه، جهت ارائه پاسخ های مناسب به ازای مشاهده داده های واقعی و خارج از مجموعه داده های آموزشی پس از پایان یافتن آموزش شبکه، کاسته می گردد. بدین ترتیب و برخلاف توقع طراح به خاطر سپاری جایگزین یادگیری روند تولید داده ها می شود.

به روشنی می توان تشابهی مابین آموزش شبکه عصبی پس انتشار خطا و عمل برازش منحنی[۳] ترسیم نمود. با استفاده از توابع چند جمله ای می توان عمل برازش منحنی بر روی تعدادی نقاط موجود از یک تابع را انجام داد. دقت در این عمل می تواند با بالاتر رفتن تعداد درجات آزادی چند جمله ای[۴] که به صورت ضرایب مقادیر توانی در چند جمله ای ها ظاهر می شوند، افزایش یابد تا جایی که تابع به دست آمده توانایی بازنمایی تمامی نقاط نمونه به طور کامل و دقیق را دارا باشد. اما چنینی تابعی مطمئناً در بازنمایی مقادیر جدیدی از تابع هدف دچار مشکل خواهد گشت، خصوصاً هنگامی که داده های نمونه موجود همچون اکثر مسائل حقیقی حاوی نوفه نیز باشند. به عبارت دیگر تقریب یک تابع با تعداد محدود داده های موجود می تواند توسط یک تابع چند جمله ای با حداقل تعداد درجات آزادی برابر با تعداد داده های موجود، به طور کامل انجام گیرد. اما استفاده از چنین تابع تقریبی باعث بروز پدیده برازش اضافی می گردد. از طرف دیگر انتخاب درجات آزادی، کمتر از مقدار مورد نیاز نیز باعث عدم توانایی مناسب جهت تقریب تابع هدف با استفاده از داده های موجود می گردد [۴].

بنابراین جهت رسیدن به یک بازنمایی مناسب همراه با قدرت تعمیم بالا، تعداد درجات آزادی و یا به صورت کلی مقدار پیچیدگی طرح و قالب مورد استفاده بسیار حائز اهمیت می باشد.

مسئله یافتن مناسب ترین پیچیدگی برای یک طرح، مثالی از قاعده ای موسوم به تیغ اوکام[۵] است، که به صورت زیر بیان می گردد.

در صورتی که چندین نظریه برای یک پدیده وجود داشته باشد، بهتر آن است که ساده ترین آنها، که به اندازه کافی جوابگوی داده های موجود باشد را انتخاب نمود.

بنابراین قانون می بایست، ساده ترین طرح ها در مقابل طرح های پیچیده انتخاب گردند، به گونه ای که مدل ساده انتخاب شده قادر به بازنمایی مناسب داده های موجود باشد.

علاوه بر مسئله برازش اضافی، کاهش درجات آزادی، در افزایش سرعت یادگیری و همچنین کاهش احتمالی تعداد نقاط کمینه محلی برای تابع معیار آموزش، که مسئله مهمی در یادگیری شبکه های عصبی می باشد، نیز می تواند مؤثر باشد. به ویژه اگر کاهش وزنها به حذف تعدادی از اتصالات نیز منجر شود. به این ترتیب شبکه از حالت تقارن که برای یادگیری، مشکل آفرین است، خارج می گردد.

یک راه حل مناسب جهت افزایش توانایی تعمیم یک طرح، کنترل پیچیدگی مؤثر آن است. پیچیدگی در شبکه را می توان توسط تعداد وزنها و یا مقدار وزنها سنجید. بنابر این جهت کاهش پیچیدگی می بایست رویکرد بهینه سازی معماری را به کار بست. معماری شبکه را می توان به طور مستقیم (روش های هرس [۵] [۶] [۷]) و یا غیر مستقیم (خاتمه زود هنگام[۶] [۸] و یا انتظام [۹]) انجام داد.

بخش ۱ از این مقاله به شرح انتظام در شبکه های رو به جلو می پردازد و در بخش ۳ برخی از روش های موجود جهت انتخاب مقدار مناسب برای ضرایب انتظام

بررسی می‌گردند. در بخش ۳ روشی جهت تغییر مقادیر ضرایب انتظام متعدد در حین آموزش و با استفاده از داده‌های آموزشی معرفی گردیده و در بخش ۴ به توضیح الگوریتم استفاده شده جهت آموزش در شبیه سازی‌ها می‌پردازیم. دربخش ۵ نیز ضمن معرفی مسئله مورد نظر، نتایج به دست آمده از شبیه سازی ها و مقایسه این نتایج با نتایج گزارش شده در دیگر مقالات ارائه می‌گردد. نهایتاً در بخش ۶ اقدام به نتیجه گیری می‌نماییم.

۱ ـ انتظام در شبکه‌های رو به جلو

نیاز به انتظام می‌تواند به دو دلیل زیر باشد [۱۰]: نخست آنکه، این عمل موجب رفع مشکلات عددی در فرآیند آموزش توسط هموارسازی[۷] تابع معیار با معرفی انحناء و خمیدگی[۸] در مناطقی از سطح خطا که انحنایی کم و حتی نزدیک به صفر دارند، می‌گردد. دوم آنکه، انتظام ابزاریست جهت کاهش پراش[۹] توسط معرفی پیشقدر[۱۰] اضافی.

کاهش پیچیدگی در شبکه با استفاده از عبارت انتظام غالباً با اضافه نمودن عبارتی که نمایشی از میزان پیچیدگی شبکه می‌باشد (به عنوان مثال تعداد وزن‌های موجود در شبکه) به تابع خطا انجام می‌پذیرد، مثلاً $C_{(w)} = S_{(w)} + \lambda R_{(w)}$ ، که در آن $C_{(w)}$ تابع معیار، $S_{(w)}$ تابع خطا (به عنوان مثال مجموع مربعات خطا)، $R_{(w)}$ عبارت انتظام(به عنوان مثال مجموع توان دوم مقادیر وزن‌ها) و λ ضریب انتظام که تعیین کننده تأثیر عبارت انتظام در مقابل خطای حاصل از شبکه در طول کمینه سازی تابع معیار موردنظر می‌باشند. به اینگونه توابع معیار به عنوان معیار کمترین طول توصیف[۱۱] نیز ارجاع می‌گردد [۱۱] [۱۲] [۴].

در این صورت مقداری مناسب برای ضریب انتظام، λ، می‌توان یافت که خطای تعمیم شبکه با استفاده از عبارت انتظام بهتر از شبکه بدون استفاده از عبارت انتظام و یا حداقل به ازای $\lambda_{opt}=0$ برابر با آن گردد. بنابراین یکی از تلاش‌های مناسب، سعی در یافتن مقدار بهینه ضریب انتظام می‌باشد.

براساس عبارات مختلف ارائه شده به عنوان عبارت انتظام، روش‌های متفاوتی جهت کنترل پیچیدگی در شبکه‌های عصبی معرفی گشته‌اند. روش‌هایی چون تنزل وزن[۱۲] [۱۳] [۱۴] [۱۵] با عبارت انتظام به شکل Σw_i^2 و حذف وزن[۱۳] [۱۶] [۱۷] [۱۸] با عبارت انتظام به شکل $\Sigma w_i^2 \ (w_0^2+w_i^2)^{-1}$ که در آن w_0 مقیاسی است جهت اعمال ترجیح تعداد وزن‌های بیشتر با مقادیر کوچکتر (مقادیر بزرگ w_0) یا تعداد وزن‌های کمتر با مقادیر بزرگتر (مقادیرکوچک w_0)، از این قبیل می‌باشند.

نوع تغییر یافته تنزل وزن ، تنزل دسته‌ای وزن[۱۴] می‌باشد که توسط تعیین یک عبارت پیشقدر به صورت $w_k (1+\lambda w_k)^{-1}$ یا $(1- e^{-\lambda w}k)$ که در آنها $w_k = \sum_j |w_{jk}|$ می‌باشد، عمل تغییر و تنزل وزن‌ها برای گروهی از وزنها (به عنوان مثال وزنهای ورودی به یک گره) به طور مشترک انجام می‌گیرد [۱۳] [۱۹].

نوع گسترش یافته انتظام، انتصاب یک ضریب انتظام مجزا برای هر وزن می‌باشد. بدین ترتیب هر وزن می‌تواند بسته به وضعیت خود و میزان تأثیری که در نتیجه آموزش از خود نشان می‌دهد، ضریب انتظام خاص خود را داشته باشد. حتی در این حالت مقدار گیری برخی از ضرایب انتظام وزنها به یک عدد نسبتاً بزرگ می‌تواند موجب حذف وزن مربوطه از معماری شبکه و وقوع رویداد هرس گردد.

۲ ـ انتخاب مقادیر مناسب برای ضرایب انتظام

تعداد ضرایب انتظام به کار گرفته شده در مقالات متفاوت به گونه‌های مختلف بوده است. در بسیاری از مقالات تنها از یک ضریب انتظام واحد، جهت کنترل پیچیدگی کل شبکه استفاده گردیده است. برخی دیگر از مقالات وزنها را در گروه‌های متفاوت و بنا به پیش فرض‌های مختلف دسته بندی کرده و برای هر دسته از وزنها یک ضریب انتظام مجزا در نظر می‌گیرند. به عنوان نمونه کل وزن‌های یک شبکه را می‌توان جزو دو دسته وزنهای ورودی ـ پنهان و پنهان ـ خروجی و یا جزو چهار دسته وزنهای ورودی ـ پنهان، پیشقدر پنهان، پنهان ـ خروجی و پیشقدر خروجی فرض نمود [۱۰] [۲۲]. در برخی دیگر از مقالات نیز تعداد ضرایب انتظام جهت کنترل پیچیدگی تا حد یک ضریب انتظام به ازای هر وزن افزایش می‌یابد [۲۳] [۲۴].

در تمامی این گونه‌ها یکی از مشکلات اساسی یافتن مقادیر مناسب برای ضرایب انتظام می‌باشد. بدیهی است با بالاتر رفتن تعداد ضرایب انتظام، انتظار کنترل مناسبتری به لحاظ پیچیدگی شبکه می‌رود، اما در این صورت مشکل دیگری را وارد مسئله می‌نماییم و آن تعیین مقادیر مناسب ضرایب انتظام به کار گرفته شده

می باشد.

ابتدایی ترین روش جهت مقداردهی به شکل سعی و خطا می باشد. در این صورت می بایست شبکه را با مقادیر مختلف ضرایب انتظام به دفعات آموزش داده و باتوجه به بهترین پاسخ، مقدار مناسب ضرایب انتظام را تعیین کنیم. در این روش به لحاظ وجود فضای جستجوی وسیع، تعداد دفعات مورد نیاز برای آموزش بالا می باشد. همچنین با بالاتر رفتن تعداد ضرایب انتظام، فضای جستجو نیز به طور نمایی افزایش یافته و در واقع یافتن مقادیر بهینه برای ضرایب انتظام را غیرعملی می نماید. لذا، نیاز به روش هایی چه به طور نظری و چه به طور عملی جهت یافتن مقادیر بهینه ضرایب انتظام، احساس می گردد.

در [۲۰] [۲۱] روشی جهت یافتن مقدار بهینه ضریب انتظام به طور نظری[15] و بر حسب تعداد داده های آموزشی و تعداد وزن های تطبیق پذیر و با استفاده از مکانیک آماری ارائه گشته است. در این مقالات تنها یک ضریب انتظام برای کل شبکه در نظر گرفته شده و شبکه مورد نظر نیز یک پرسپترون بدون لایه پنهان می باشد. بنابراین، روش مذکور جامعیت خود را در رابطه با شبکه های چند لایه از دست داده و بسط به شبکه های غیرخطی به عنوان یک مسئله باز مطرح می گردد.

در [۱۸] نیز رویه ای جهت تغییر مقدار ضریب انتظام، در حین آموزش برای روش حذف وزن ارائه گشته است. در این روش نیز تنها از یک ضریب انتظام برای کل شبکه استفاده شده است.

در [۲۲] دو ضریب انتظام مجزا برای وزن های ورودی - پنهان و پنهان - خروجی در نظر گرفته شده است. جهت یافتن مقدار بهینه آنها، فضای مقادیر برای ضرایب انتظام را به شکل یک توری[16] 3×3 تقسیم بندی کرده و مقدار خطای شبکه در نقاط گره توری، به دست آورده شده است. سپس یک سطح سهموی به نقاط خطا خورانده شده ونقطه کمینه سطح سهموی را به دست آورده اند. نهایتاً از مقادیر موجود برای ضرایب انتظام در نقطه کمینه به عنوان مقادیر بهینه ضرایب انتظام استفاده گشته است.

در [۲۳] [۲۴] خطای اعتبارسنجی[17] به طور ضمنی تابعی از ضرایب انتظام معرفی شده و مقادیر ضرایب انتظام براساس حرکت بر خلاف جهت بردار گرادیان تابع خطای اعتبارسنجی نسبت به ضرایب انتظام تغییر داده می شوند. در این رویه ابتدا شبکه با مقادیر ثابت برای ضرایب انتظام، تا رسیدن به نقطه کمینه، آموزش داده شده و سپس ضرایب انتظام، در جهت خلاف بردار گرادیان خطای اعتبارسنجی نسبت به ضرایب انتظام به گونه ای تغییر می یابد که خطای اعتبارسنجی کمینه گردد. در انتها نیز شبکه، با مقادیر یافته شده برای ضرایب انتظام و با استفاده از تمامی داده های موجود آموزش داده می شود. در این روش نیازمند تقسیم داده های موجود به دو دسته آموزشی و اعتبارسنجی می باشیم. بنابر این هم رویه تغییر وزنها و هم رویه تغییر مقادیر ضرایب انتظام براساس تعداد داده های کمتری صورت می پذیرد، که می تواند باعث افت کیفیت در آموزش باشد. چگونگی تقسیم بندی و میزان تقسیم بندی نیز خود می تواند به عنوان مشکلی در این روش قابل طرح باشد.

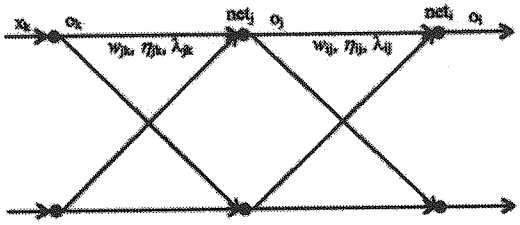
در [۲۵] همان روش برای یک شبکه دسته بندی کننده SOFTMAX استفاده شده است.

کلیات روش در [۱۰] همانند [۲۳] [۲۴] و [۲۵] بوده با این تفاوت که در [۱۰] از بردار گرادیان خطای اعتبارسنجی متقابل[18] جهت تغییر مقادیر ضرایب انتظام استفاده شده است.

در این مقاله مقادیر ضرایب انتظام براساس کاهش خطای حاصل از شبکه تغییر داده شده و همزمان با حرکت بر روی سطح تابع معیار جهت بهینه سازی وزنها، حرکت بر روی سطح خطا جهت بهینه سازی مقادیر ضرایب انتظام نیز صورت می پذیرد. در روش ارائه شده از تمامی داده های موجود جهت بهینه سازی وزنها و ضرایب انتظام استفاده می گردد. نتایج حاصل از آزمایشات نشان می دهد که به دلیل تطبیق ضرایب انتظام به شکل تکراری و در طول فرایند آموزش، نیاز به بازآموزی مجدد شبکه با مقادیر یافته شده برای ضرایب انتظام به صورت مقادیری ثابت نمی باشد. این نتایج در بخش ۵ ارائه گردیده اند.

۳ - تغییر مقادیر ضرایب انتظام براساس کاهش میزان خطا در خروجی (با مشاهده بردار گرادیان)

معماری شبکه رو به جلوی چند لایه ای را که در آن هر یال (سیناپس) دارای مشخصه های مقدار وزن w، طول گام η و ضریب انتظام λ، مختص به خود می باشد را در نظر می گیریم.



شکل (۱) شمای جامعی از یك شبکه رو به جلو.

که روابط حاکم بر آن را می توان به شکل زیر نوشت،

$$o_i = f(net_i) \qquad (۱)$$

$$o_j = f(net_j) \qquad (۲)$$

$$net_i = \sum_j o_j (w_{ij} + \Delta w_{ij}) \qquad (۳)$$

$$net_j = \sum_k o_k (w_{jk} + \Delta w_{jk}) \qquad (۴)$$

جهت آموزش شبکه اقدام به تغییر وزن ها در خلاف جهت بردار گرادیان نسبت به وزن ها به منظور کمینه سازی تابع معیار (۵) می نماییم.

$$C = S + R$$

$$C = \frac{1}{2}\sum_i (o_i - t_i)^2 + \frac{1}{2}\sum_{u \in W} \lambda_u w_u^2 \qquad (۵)$$

که در آن S خطای حاصل از شبکه به صورت مجموع مربعات خطا، R عبارت انتظام به منظور کنترل پیچیدگی شبکه و W مجموعه تمامی وزن های شبکه می باشند. باتوجه به تابع معیار (۵)، مقدار تغییر وزن ها در هر تکرار به طریق زیر محاسبه می گردد.

$$\Delta w_{ij} = -\eta_{ij} \frac{\partial C}{\partial w_{ij}} = -\eta_{ij}\left(\frac{\partial S}{\partial w_{ij}} + \frac{\partial R}{\partial w_{ij}}\right)$$

$$\Delta w_{ij} = -\eta_{ij}\left(\frac{\partial S}{\partial w_{ij}} + \lambda_{ij} w_{ij}\right) \qquad (۶)$$

$$\Delta w_{jk} = -\eta_{jk} \frac{\partial C}{\partial w_{jk}} = -\eta_{jk}\left(\frac{\partial S}{\partial w_{jk}} + \frac{\partial R}{\partial w_{jk}}\right)$$

$$\Delta w_{jk} = -\eta_{jk}\left(\frac{\partial S}{\partial w_{jk}} + \lambda_{jk} w_{jk}\right) \qquad (۷)$$

جهت تغییر مقادیر λ_{ij}ها رویه ای مشابه به رویه یادگیری وزن های بهینه برای کمینه سازی تابع معیار (۵) انجام می دهیم. در صورتی که هدف به دست آوردن میزان تغییرات در ضرایب انتظام براساس کمینه سازی این تابع معیار باشد، به سادگی می توان دریافت که با توجه به شکل خاص آن، کمینه سازی سیری نزولی را برای λ_{ij}ها رقم خواهد زد و هر چه این کاهش مقدار λ_{ij} ها بیشتر انجام گردد، کمینه سازی تابع معیار (۵) موفق تر خواهد بود.

بنابراین در صورتی که محدودیتی برای مقادیر λ_{ij}ها در نظر گرفته نشود، می توان انتظار داشت که مقدار عددی آنها تمایل به میل به سمت $-\infty$ داشته باشند و در صورت در نظر گرفتن محدودیت $\lambda_{ij} \geq 0$ ، بدیهی است تمامی λ_{ij}ها به سمت صفر حرکت کرده و در آنجا ثابت می مانند. در این صورت آموزش تبدیل به کمینه سازی خطا بدون استفاده از عبارت انتظام می گردد.

برای رفع این نقیصه از تابع معیار دیگری برای بهینه سازی مقادیر ضرایب انتظام استفاده می نماییم. با ملاحظه دقیق تابع S می توان دریافت که به دلیل استفاده از λ_{ij} در تعیین مقدار فعلی وزن با کمک مقدار قبلی آن، وابستگی به λ در مقدار فعلی ایجاد می شود، لذا تابع S می تواند تابعی از λ_{ij} تلقی گردد. این نکته را می توان با استفاده از معادلات (۶) و (۷) ملاحظه نمود. بنابراین سعی در بهینه سازی مقادیر ضرایب انتظام توسط رویه تکراری پس انتشار خطا، جهت کمینه سازی تابع خطای S را داریم. از این رو حرکت بر روی فوق سطح خطا، و در خلاف جهت بردار گرادیان نسبت به ضرایب انتظام، به طریق زیر انجام می پذیرد.

$$S = \frac{1}{2}\sum_i (o_i - t_i)^2 \qquad (۸)$$

$$\lambda_{ij} = \lambda_{ij} + \Delta\lambda_{ij} \qquad (۹)$$

$$\Delta\lambda_{ij} = -\mu \frac{\partial S}{\partial \lambda_{ij}} \qquad (۱۰)$$

که در آن μ طول گام جهت تغییر مقادیر ضرایب انتظام می باشد. برای محاسبه گرادیان تابع خطای S نسبت به ضرایب انتظام، با استفاده از قاعده زنجیره ای به طریق زیر عمل می نماییم.

$$\frac{\partial S}{\partial \lambda_{ij}} = \frac{\partial S}{\partial o_i}\frac{\partial o_i}{\partial net_i}\frac{\partial net_i}{\partial \lambda_{ij}} \qquad (۱۱)$$

$$\frac{\partial net_i}{\partial \lambda_{ij}} = \frac{\partial}{\partial \lambda_{ij}}\left[\sum_j o_j\,(w_{ij} + \Delta w_{ij})\right]$$

$$\frac{\partial net_i}{\partial \lambda_{ij}} = \frac{\partial}{\partial \lambda_{ij}}\left[\sum_j o_j\,\Delta w_{ij}\right] \qquad (۱۲)$$

با جایگزینی (۶) در (۱۲) خواهیم داشت،

$$\frac{\partial net_i}{\partial \lambda_{ij}} = \frac{\partial}{\partial \lambda_{ij}}\left[\sum_j -\eta_{ij}\,o_j\left(\frac{\partial S}{\partial w_{ij}} + \lambda_{ij}\,w_{ij}\right)\right]$$

$$\frac{\partial net_i}{\partial \lambda_{ij}} = -\eta_{ij}\,o_j\,w_{ij} \qquad (۱۳)$$

بدین ترتیب و با جایگزینی (۱۳) در (۱۱) خواهیم داشت،

$$\frac{\partial S}{\partial \lambda_{ij}} = \frac{\partial S}{\partial o_i}\,\frac{\partial o_i}{\partial net_i}\,o_j\,(-\eta_{ij}\,w_{ij}) \qquad (۱۴)$$

و نهایتاً گرادیان تابع خطای S نسبت به ضرایب انتظام به شکل ساده زیر قابل تحریر می‌باشد.

$$\frac{\partial S}{\partial \lambda_{ij}} = -\eta_{ij}\,w_{ij}\,\frac{\partial S}{\partial w_{ij}} \qquad (۱۵)$$

به طریق مشابه می‌توان این رابطه را برای وزن‌های موجود در لایه‌های پایین‌تر شبکه نیز به دست آورد.

$$\frac{\partial S}{\partial \lambda_{jk}} = -\eta_{jk}\,w_{jk}\,\frac{\partial S}{\partial w_{jk}} \qquad (۱۶)$$

همانگونه که از شکل روابط (۱۵) و (۱۶) پیداست، محاسبه گرادیان خطا نسبت به ضرایب انتظام بار محاسباتی ناچیزی را به رویه معمول آموزش تحمیل می‌نماید، چرا که به طور معمول ناگزیر به محاسبه گرادیان خطا نسبت به وزنها، $\partial S/\partial w_{ij}$، می‌باشیم.

٤ ـ الگوریتم آموزش

الگوریتم استفاده شده در این مقاله برای آموزش شبکه به شکل دسته‌ای [۱۹] و بر مبنای تقریب محلی از درجه دوم تابع معیار (۵) می‌باشد[۲۰].

با استفاده از تقریب محلی درجه دوم، می‌توان مستقیماً عبارتی، جهت مشخص نمودن محل کمینه تابع معیار به شکل زیر به دست آورد،

$$w^* = w - H^{-1}\,g \Rightarrow \Delta w = -H^{-1}\,g \qquad (۱۷)$$

که در آن g بردار گرادیان و H ماتریس هسیین[۲۱] می‌باشند و به طریق زیر محاسبه می‌گردند،

$$g = \nabla C_{(w)}\,|_{w=w_0} = \left.\frac{\partial C_{(w)}}{\partial w}\right|_{w=w_0} \qquad (۱۸)$$

$$H = \nabla\nabla C_{(w)}\,|_{w=w_0} = \left.\frac{\partial^2 C_{(w)}}{\partial w^T\,\partial w}\right|_{w=w_0} \qquad (۱۹)$$

روش استفاده شده موسوم به روش نیوتن بوده و بردار $H^{-1}g$ ـ به عنوان جهت نیوتن و یا گام نیوتن شناخته میشود [۴] که مبنای بسیاری از راه کارهای بهینه سازی می‌باشد.

با بالا رفتن تعداد وزن‌های موجود در شبکه، اندازه ماتریس H شدیداً افزایش یافته و محاسبه مقادیر عناصر ماتریس H و همچنین معکوس این ماتریس بسیار زمان‌گیر می‌گردد، به طوری که تقریباً محاسبه آن را غیر عملی می‌نماید. یک راه حل ساده استفاده تقریبی از ماتریس H، موسوم به ماتریس شبه هسیین[۲۲]، $\widetilde{H}$، به وسیله حذف عناصر غیر قطری ماتریس H می‌باشد [۲۶]. به این ترتیب نسخه‌ای تقریبی از روش نیوتن به نام روش شبه نیوتن[۲۳] معرفی می‌گردد. استفاده از این تقریب علاوه بر عملی ساختن محاسبه ماتریس $\widetilde{H}$، به دست آوردن معکوس آن را نیز نیازمند محاسبات ناچیزی می‌گرداند. به علاوه استفاده از این تقریب، روابط بروزرسانی وزن‌ها را نیز به شکل روابط مجزایی برای هر وزن تبدیل می‌نماید.

$$\Delta w_{ij} = -\left(\frac{\partial^2 C}{\partial w_{ij}^2}\right)^{-1}\frac{\partial C}{\partial w_{ij}} \qquad (۲۰)$$

در برخی نواحی از فضای وزن‌ها که سطح خطا دارای انحناء و خمیدگی کمی باشد، مقدار مخرج عددی نزدیک به صفر شده و لذا علاوه بر بروز مشکل در محاسبات، باعث جهش‌های بزرگ و ناخواسته در مقادیر وزن‌ها می‌گردد. جهت جلوگیری از بروز چنین مشکلی، مخرج کسر رابطه (۲۰) را با عددی ثابت، ۷، و نسبتاً کوچک جمع می‌نماییم.

به علاوه در صورتی که مقدار موجود در مخرج رابطه (۲۰) عددی منفی باشد، حرکت به سمت بیشینه فوق سطح خطا را منجر می‌گردد. تقریب ضرب

خارجی[۲۴] می‌تواند راه حلی برای این مشکل باشد [۴] که علاوه بر آن، محاسبات لازم برای به دست آوردن مقدار عناصر ماتریس شبه هسیین را نیز تا حد زیادی کاهش می‌دهد. اما شرایط لازم برای معتبر شمردن چنین تقریبی همواره برقرار نمی‌باشد، بنابراین در این مقاله مقادیر دقیق عناصر ماتریس شبه هسیین محاسبه گشته [۴] و قدر مطلق این مقادیر مورد استفاده قرار گرفته اند.

اما تقریب درجه دوم از تابع معیار، تنها هنگامی معتبر است که $w_{ij(new)}$ به اندازه کافی نزدیک به $w_{ij(old)}$ باشد، در صورتی که ممکن است مقدار به دست آمده برای تغییر وزن ها، منجر به تخلف از تقریب درجه دوم گردد. بنابر این، معرفی یک ضریب گام $\beta > 0$ جهت تضمین اینکه مقدار تغییر وزن ها نهایتاً منجر به کاهش خطا گردد، سودمند می‌باشد. بدین ترتیب رابطه نهایی جهت تغییر وزن ها برابر خواهد بود با،

$$\Delta w_{ij} = -\beta \left(\left| \frac{\partial^2 C}{\partial w_{ij}^2} \right| + \gamma \right)^{-1} \frac{\partial C}{\partial w_{ij}} \qquad (۲۱)$$

در هر تکرار β مقدار اولیه ای برابر با عدد ۱ داشته و تا جایی که تغییر در وزن ها منجر به کاهش مقدار تابع معیار گردد بر دو تقسیم می‌شود.

نهایتاً الگوریتم مورد استفاده براساس آنچه در [۲۷] بیان شده و با احتساب تغییرات مورد نظر جهت تغییر مقادیر ضرایب انتظام به شکل زیر می‌باشد.

مرحله ۱ ـ مقدار گذاری اولیه

۱ ـ ۱ ـ وزنها را توسط بردار w_0 مقدار گذاری اولیه نموده و مقدار تابع معیار، $C_{N(w0)}$، را محاسبه می‌کنیم.

۱ ـ ۲ ـ مقدار قدیمی تابع معیار را برابر با بی نهایت فرض می‌کنیم[۲۵]. $C_{N(w(c-1))} = \infty$

۱ ـ ۳ ـ مقداری کوچک به عنوان حد آستانه، $1 >> \tau$، در نظر می‌گیریم. این حد آستانه مشخص کننده کوچکترین تغییر نسبی مورد قبول در مقدار تابع معیار می‌باشد.

۱ ـ ۴ ـ بیشترین تعداد تکرار، itr، را تعیین می نماییم.

۱ ـ ۵ ـ ضریب گام را با بزرگترین مقدار مجاز خود مقدار گذاری اولیه می نماییم. $\beta = 1$

۱ ـ ۶ ـ ضرایب انتظام را مقدار گذاری اولیه می نماییم.

۱ ـ ۷ ـ شمارنده c را مقدارگذاری اولیه می نماییم. $c = 0$

مرحله ۲

$$\text{while } (c < itr) \text{ and } \left(\frac{C_{N(w(c-1))} - C_{N(w(c))}}{C_{N(w(c-1))}} > \tau \right)$$

۲ ـ ۱ ـ $c \leftarrow c + 1$

۲ ـ ۲ ـ محاسبه بردار گرادیان نسبت به وزن ها

۲ ـ ۳ ـ محاسبه ماتریس شبه هسیین

۲ ـ ۴ ـ $C_{N(w(c))} = \infty$ و $\beta = 1$

۲ ـ ۵ ـ محاسبه میزان تغییر در وزن ها براساس رابطه (۲۱)

۲ ـ ۶ ـ while $(C_{N(w(c-1))} < C_{N(w(c))})$

۲ ـ ۶ ـ ۱ ـ محاسبه مقادیر $\eta_{ij} = \beta \left(\left| \frac{\partial^2 C}{\partial w_{ij}^2} \right| + \gamma \right)^{-1}$ برای تمامی وزن ها

۲ ـ ۶ ـ ۲ ـ نگهداری مقادیر $\eta_{ij} w_{ij}$ برای تمامی وزنها جهت استفاده در مرحله ۲–۸

۲ ـ ۶ ـ ۳ ـ $w_{(c)} \leftarrow w_{(c-1)} + \beta \Delta w$

۲ ـ ۶ ـ ۴ ـ محاسبه $C_{N(w(c))}$

۲ ـ ۶ ـ ۵ ـ $\beta \leftarrow \beta/2$

۲ ـ ۷ ـ پایان حلقه ۲ ـ ۶

۲ ـ ۸ ـ محاسبه بردار گرادیان نسبت به ضرایب انتظام بر اساس روابط (۱۵) و (۱۶)

۲ ـ ۹ ـ تغییر مقادیر ضرایب انتظام براساس روابط (۹) و (۱۰)

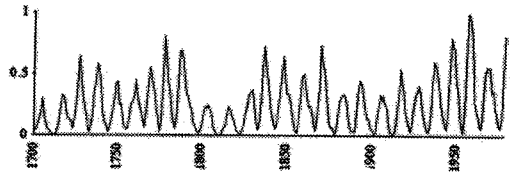
پایان حلقه مرحله ۲

۵ ـ نتایج تجربی

جهت بررسی کارآیی الگوریتم ارائه شده و توانایی مقایسه نتایج به دست آمده از آن با دیگر روش های به کار گرفته شده، استفاده از داده های مربوط به لکه های خورشیدی انتخاب گردیده است.

این داده ها پس از ارائه در مقاله [۲۸] به عنوان یک سکوی مقایسه ای در ارائه عملکرد شبکه های عصبی مورد استفاده قرار گرفت. از آنجا که ذات مسئله مربوط به علم نجوم می‌باشد، از چند و چون و تعبیر مفاهیم داده ها صرفنظر کرده و صرفاً از آن به عنوان داده های یک مسئله واقعی با ماهیت سری های زمانی و با رفتار آشوبگونه یاد می نماییم. داده های موجود که حاوی اطلاعات مقیاس شده سالیانه، مربوط به لکه های خورشیدی از سال ۱۷۰۰ تا ۱۹۷۹ میلادی می‌باشد، به سه دسته مجزا تقسیم می‌گردند. یک دسته جهت آموزش، حاوی داده های مربوط به سال های ۱۷۰۰ تا

۱۹۲۰ میلادی و دو دسته جهت آزمایش، حاوی داده های مربوطه به سال های ۱۹۲۱ تا ۱۹۵۵ و ۱۹۵۶ تا ۱۹۷۹ میلادی. بدین ترتیب داده های دسته آزمایشی اول در مقایسه با داده های دسته آزمایشی دوم، از قرابت زمانی بیشتری به داده های دسته آموزشی برخوردار می باشند. با توجه به شکل ۲ الگوهایی با بیشینه های محلی، به صورت چرخشی قابل ملاحظه می باشند. زمان میان این محدوده از ۷ تا ۱۷ سال با میانه ۱۱ سال است.



شکل (۲) روند تغییرات داده های لکه های خورشیدی از سال ۱۷۰۰ تا ۱۹۷۹ میلادی.

بنابر این می توان هدف، جهت شناسایی را به این ترتیب معرفی نمود که، می خواهیم مقدار مقیاس شده در زمان ۱۲ را با استفاده از مقادیر مقیاس شده مربوطه به ۱۱ زمان قبل، تقریب بزنیم. جهت هم خوانی با دیگر مقالات، شبکه بکار گرفته شده، یک شبکه ۱ × ۸ × ۱۲ با توابع (.) tanh برای گره های پنهان و تابع خطی برای گره خروجی بوده و خطای مقیاس شده شبکه پس از آموزش، به طریق زیر محاسبه می گردد.

$$E_{set} = \frac{1}{\sigma^2_{total}} \frac{1}{P_{set}} \sum_{m=1}^{P_{set}} (o^{(m)} - t^{(m)})^2 \qquad (۲۲)$$

که در آن E_{set} خطای حاصل از هر دسته مورد بررسی، σ^2_{total} پراش موجود در کل داده ها[۲۶] و P_{set} تعداد داده های موجود در هر دسته مورد بررسی[۲۷] می باشند.

شبیه سازی به سه روش برای دوازده نقطه شروع انجام و تعداد نه عدد از بهترین جواب ها انتخاب گردیده اند. سه روش مورد استفاده به شرح زیر می باشند،

الف ـ آموزش شبکه توسط الگوریتم شبه نیوتن و بدون عبارت انتظام

ب ـ آموزش شبکه توسط الگوریتم شبه نیوتن با عبارت انتظام تنزل وزن متداول با استفاده از یک ضریب انتظام واحد و ثابت در طول آموزش

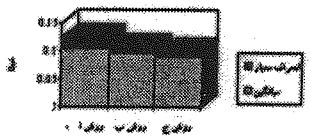
ج ـ آموزش شبکه توسط الگوریتم شبه نیوتن با عبارت انتظام تنزل وزن با استفاده از یک ضریب انتظام تطبیق پذیر در حین آموزش برای هر وزن (روش جدید ارائه شده در این مقاله)

انتخاب مقدار ضریب انتظام ، λ، در روش ب و طول گام تغییر ضرایب انتظام، μ، در روش ج به طور سعی و خطا انجام گردیده است. هر دو روش به ازای هشت مقدار مختلف برای این مقادیر شبیه سازی گردیده اند.[۲۸] همانگونه که ملاحظه می گردد انتخاب عدد مناسب برای طول گام ضریب انتظام تعداد زیادی شبیه سازی را می طلبد. انتخاب خودکار این عدد خود می تواند موضوعی برای کارهای آتی باشد.

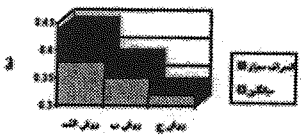
در روش ارائه شده (روش ج) کلیه ضرایب انتظام با مقدار ۰/۰۰۰۱ مقدار گذاری اولیه گردیده اند.

جدول ۱ نشان دهنده نتایج به دست آمده از آموزش نه شبکه مورد نظر توسط سه روش عنوان شده، به همراه نتایج نهایی به صورت میانگین و انحراف معیار خطاها می باشد.

شکل ۳ و شکل ۴ نیز شمایی مقایسه ای از نتایج نهایی خطاهای دسته های آزمایشی اول و دوم برای سه روش مورد نظر را به نمایش می گذارد.



شکل (۳) نتایج نهایی خطاهای دسته آزمایشی اول برای هر سه روش.



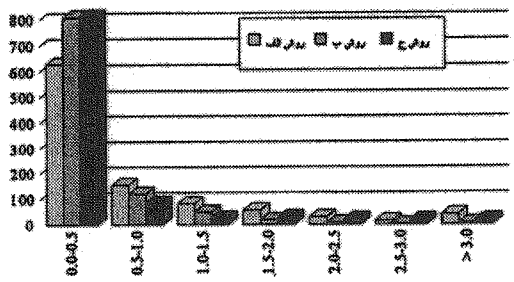
شکل (۴) نتایج نهایی خطاهای دسته آزمایشی دوم برای هر سه روش.

باتوجه به این شکل ها و با استفاده از نتایج مندرج در جدول ۱ ملاحظه می گردد که میانگین و انحراف معیار خطای داده های آزمایشی اول و دوم، از روش الف تا روش این مقاله (ج)، تماماً کاهش یافته اند.

کاهش خطاهای روش این مقاله (روش ج) نسبت به روش ب، برای هر دو مجموعه آزمایشی اول و دوم، نمایانگر عملکرد بهتر شبکه در صورت استفاده از ضرایب انتظام متعدد تطبیق پذیر در مقابل استفاده از یک

ضریب انتظام واحد با مقدار ثابت است.

هدف در استفاده از عبارت انتظام به طور اعم، کنترل پیچیدگی شبکه و استفاده از عبارت انتظام تنزل وزن به طور اخص، کنترل پیچیدگی از طریق کوچک نگاه داشتن مقادیر وزن ها در کل شبکه می باشد. شکل ۵ نشان دهنده سابقه نمای[۲۹] قدر مطلق مقدار وزن ها، حاصل از اجتماع تمامی وزن های موجود در نه شبکه انتخاب شده در جدول ۱، برای هر سه روش می باشد.



شکل (۵) سابقه نمای قدر مطلق مقدار وزن ها بر روی اجتماع وزن های نه شبکه انتخاب شده برای هرسه روش، محور عمودی فراوانی و محور افقی دسته بندی مقداری وزن ها را نشان می دهند.

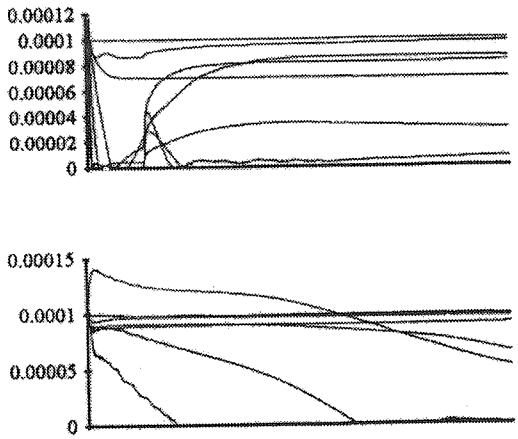
همانگونه که در این شکل مشاهده می گردد، فراوانی قدر مطلق وزن های با مقادیر بین صفر و ۰/۵ برای روش های ب و این مقاله (ج) بسیار بیش از روش الف است، که حکایت از صحت عملکرد استفاده از عبارت انتظام تنزل وزن را دارد. با بزرگتر شدن مقدار قدرمطلق وزن ها، فراوانی آنها کاهش می یابد. با توجه به این شکل مشاهده می گردد که وزن های بزرگتر از ۱/۵ در روش ب از فراوانی کمتری نسبت به روش این مقاله (ج) برخوردار هستند، در صورتی که با توجه به نتایج به دست آمده می دانیم، عملکرد روش این مقاله (ج) بهتر از روش ب بوده است. گذشته از این که تعداد وزن های موجود در این محدوده تعداد اندکی می باشند، دلیل این تفاوت را می توان حاصل از انعطاف پذیری روش این مقاله (ج) در تعیین مقادیر ضرایب انتظام مختص به هر وزن دانست. به عبارت دیگر روش ب با استفاده از یک ضریب انتظام مشترک و ثابت برای تمامی وزن ها در طول بهینه سازی، همواره تمامی وزن ها را به یک نسبت و به طور مستبدانه ملزم به کاهش می نماید، در صورتی که این عمل در روش این مقاله (ج) در صورت نیاز و به میزان لازم برای هر وزن، به طور مجزا انجام می گردد. شکل ۶ نشان دهنده روند تغییرات مقادیر ضرایب انتظام در طول بهینه سازی برای وزن های ارتباطی بین یک گره ورودی و هشت گره میانی می باشد. این شکل برای دو گره ورودی متفاوت تهیه گردیده است.

شکل ۷ سابقه نما را برای وزن های بین صفر و ۰/۵ در دسته هایی با طول ۰/۱ برای هر سه روش نشان می دهد. این شکل می تواند به نتیجه گیری در مورد صحت عملکرد روش های ب و این مقاله (ج) که فراوانی وزن های بین صفر و ۰/۵ آنها در شکل ۵ تقریباً برابربودند کمک فراوانی نماید.

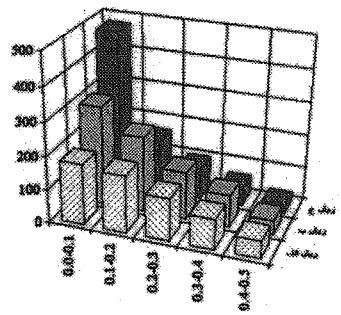
همانگونه که در این شکل ملاحظه می گردد، فراوانی وزن های بین صفر و ۰/۱ در روش این مقاله (ج)، به اندازه قابل ملاحظه ای بیش از دو روش دیگر می باشد، که دلالت بر توانایی روش مذکور در برآورد نظر طراح برای کوچک نگاه داشتن مقادیر وزن ها به جهت افزایش قدرت تعمیم شبکه عصبی دارد.

جدول ۲ نتایج حاصل از بازآموزی شبکه های انتخاب شده برای روش این مقاله (ج)، با استفاده از مقادیر نهایی ضرایب انتظام، به طور ثابت در بازآموزی را نشان می دهد. همانگونه که قبلاً نیز اشاره شد، این نتایج حکایت از عدم نیاز به بازآموزی مجدد با مقادیر یافته شده ضرایب انتظام دارند.

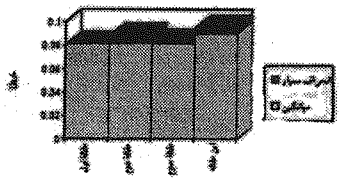
جدول ۳ نشان دهنده نتایج ارائه شده در دیگر مقالات به همراه نتیجه حاصل از روش ارائه شده می باشد. این نتایج برای دسته های آزمایشی اول و دوم به صورت مقایسه ای در شکل ۸ و شکل ۹ نمایش داده شده اند. در این شکل ها منظور از مقاله اول نتایج [۲۸]، مقاله دوم نتایج [۲۹] و مقاله سوم نتایج [۳۰] می باشد.



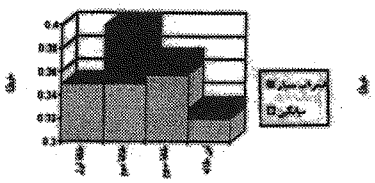
شکل (۶) روند تغییر مقدار ضرایب انتظام در طول بهینه سازی، نمودارها نشان دهنده این روند برای وزن های بین دو گره ورودی و گره های میانی می باشند.



شکل (۷) سابقه نمای قدرمطلق مقدار وزن ها برای وزن های بین صفر و ۵ / ۰ بر روی اجتماع وزن های نه شبکه انتخاب شده برای هر سه روش، محور عمودی فراوانی و محور افقی دسته بندی مقدار وزن ها را نشان می دهند.



شکل (۸) خطاهای دسته آزمایشی اول، حاصل از روش ارائه شده در این مقاله و سه مقاله دیگر.



شکل (۹) خطاهای دسته آزمایشی دوم، حاصل از روش ارائه شده در این مقاله وسه مقاله دیگر.

باتوجه به این دو شکل و نتایج مندرج در جدول ۳، مشاهده می گردد که، با آنکه خطای حاصل از داده های آزمایشی اول اندکی افزایش نشان می دهد، خطای به دست آمده از مجموعه آزمایشی دوم ، دست خوش کاهش گردیده است. توجه به مقادیر به دست آمده برای خطای مجموعه آزمایشی دوم نشان می دهد که در تمامی روش های دیگر میانگین خطای گزارش شده مقداری برابر ۰ / ۳۵ می باشد، که این مقدار توسط روش ارائه شده به ۰ / ۳۲ تقلیل یافته است. از آنجا که مجموعه داده های آزمایشی دوم نسبت به مجموعه داده های آموزشی به لحاظ زمانی دورتر از داده های آزمایشی اول می باشد، لذا بهبود خطای حاصل از آن می تواند نمایانگر حفظ مناسب تر دینامیک موجود در سیستم توسط روش مورد استفاده محسوب گردد. جریمه حفظ این دینامیک نیز افزایش اندکی است که در خطای داده های آزمایشی اول پرداخته ایم.

۶ـ نتیجه گیری

در این مقاله روشی تحلیلی جهت تعیین تطبیقی مقادیر ضرایب انتظام در حین آموزش معرفی گردید.

نتایج حاصل از شبیه سازی های مختلف مؤید افزایش قدرت تعمیم توسط روش ارائه شده می باشد. بررسی مقادیر وزن های شبکه های حاصل از شبیه سازیها، بیانگر توانایی روش ارائه شده در تأمین نظر طراح جهت کوچک نگاه داشتن مقادیر وزن ها، در مقایسه با دیگر روش ها می باشد. مقایسه این نتایج با جواب های گزارش شده در دیگر مقالات نشان دهنده حفظ بهتر دینامیک سیستم در مواجه با داده هایی، با فاصله زمانی بیشتر می باشد.

انتخاب خودکار طول گام جهت تعیین مقدار تغییر لازم در مقدار ضرایب انتظام و ادغام روش ارائه شده با روشی جهت حذف قطعی وزن های کم اثر به عنوان فعالیت های آتی معرفی می گردند.

زیرنویس ها

1 - Over fitting
2 - Noise
3 - Curve Fitting
۴ ـ تعداد پارامترهای تطبیق پذیر آن
5 - Occam/ockham's razor
6 - Early Stopping
7 - Smoothing
8 - Curvature
9 - Variance
10 - Bias
11 - Minimum Discription Length (MDL)
12 - Weight Decay
13 - Weight Elimination
14 - Weight Group Decay
۱۵ ـ تنها در حالت استفاده از عبارت تنزل وزن
16- Grid
17 - Validation error
18 - Cross - Validation error
19 - Batch
۲۰ ـ توسط بسط تیلور تابع معیار از مرتبه دوم حول نقطه w_0 از فضای وزن ها
21 - Hessian
22 - Pseudo - Hessian
23 _ Pseudo - Newton

24 _ Outer Product Approximation

که به نام تقریب Levenberg - Maruardt نیز شناخته می شود.

۲۵ ـ منظور از بینهایت عددی بسیار بزرگ وقابل نمایش در شبیه سازی ها می باشد.

۲۶ ـ برای داده های لکه های خورشیدی این مقدار برابر ۰/۰۴۱۱ می باشد.

۲۷ ـ این مقدار برای داده های آموزشی برابر ۲۰۹، داده های آزمایشی اول ۳۵ و داده های آزمایشی دوم ۲۴ می باشد.

۲۸ ـ به ازای مقادیر ۰/۰۰۰۰۰۱، ۰/۰۰۰۰۰۵، ۰/۰۰۰۰۱، ۰/۰۰۰۰۵، ۰/۰۰۰۱، ۰/۰۰۰۵، ۰/۰۰۱، ۰/۰۰۵ و m به ازای مقادیر ۰/۰۰۰۱، ۰/۰۰۰۵، ۰/۰۰۰۷، ۰/۰۰۱، ۰/۰۰۳، ۰/۰۰۵، ۰/۰۰۷، ۰/۰۱

29 - Histogram

جدول (۱) نتایج حاصل از شبیه سازی به سه روش متفاوت الف) شبه نیوتن بدون عبارت انتظام، ب ـ شبه نیوتن با عبارت انتظام تنزل وزن دارای یك ضریب انتظام واحد و ابت، ج ـ شبه نیوتن با عبارت انتظام تنزل وزن دارای ضرایب انتظام تطبیق پذیر مجزا برای هر وزن (ارائه شده در بیان مقاله).

خطاهای حاصل از روش ج				خطاهای حاصل از روش ب				خطاهای حاصل از روش الف			
دستهٔ آزمایشی۲	دستهٔ آزمایشی۱	دستهٔ آموزشی	μ	دستهٔ آزمایشی۲	دستهٔ آزمایشی۱	دستهٔ آموزشی	λ	دستهٔ آزمایشی۲	دستهٔ آزمایشی۱	دستهٔ آموزشی	
۰/۳۲	۰/۰۹۰	۰/۰۹۳	۰/۰۰۰۷	۰/۳۰	۰/۱۰۳	۰/۰۷۰	۰/۰۰۰۰۰۱	۰/۴۸	۰/۱۱۵	۰/۰۷۴	
۰/۳۳	۰/۱۰۴	۰/۰۷۲	۰/۰۰۳	۰/۳۸	۰/۱۱۷	۰/۰۷۲	۰/۰۰۰۰۱	۰/۳۱	۰/۱۰۴	۰/۰۷۰	
۰/۳۱	۰/۰۹۶	۰/۰۷۳	۰/۰۰۰۵	۰/۳۵	۰/۱۰۰	۰/۰۹۹	۰/۰۰۰۰۶	۰/۴۱	۰/۱۵۷	۰/۰۵۰	
۰/۳۴	۰/۰۹۳	۰/۰۸۵	۰/۰۰۰۵	۰/۳۷	۰/۱۱۶	۰/۱۱۸	۰/۰۰۰۰۵	۰/۳۳	۰/۰۸۷	۰/۰۶۹	
۰/۳۲	۰/۰۸۶	۰/۰۸۱	۰/۰۰۰۵	۰/۳۱	۰/۰۷۸	۰/۰۸۱	۰/۰۰۰۰۵	۰/۴۸	۰/۱۰۷	۰/۰۷۴	
۰/۳۱	۰/۰۹۰	۰/۰۷۶	۰/۰۰۳	۰/۳۸	۰/۰۹۸	۰/۰۷۶	۰/۰۰۰۰۰۵	۰/۳۷	۰/۰۹۶	۰/۰۷۸	
۰/۳۲	۰/۰۸۳	۰/۰۷۹	۰/۰۰۰۱	۰/۳۷	۰/۰۹۱	۰/۰۷۶	۰/۰۰۰۰۱	۰/۴۰	۰/۱۰۹	۰/۰۷۰	
۰/۳۳	۰/۰۸۸	۰/۰۷۸	۰/۰۰۰۵	۰/۳۳	۰/۰۸۸	۰/۰۷۱	۰/۰۰۰۰۰۵	۰/۳۴	۰/۰۸۵	۰/۰۷۸	
۰/۳۳	۰/۰۸۴	۰/۰۸۷	۰/۰۰۰۱	۰/۳۷	۰/۰۹۰	۰/۰۸۷	۰/۰۰۰۰۰۵	۰/۳۶	۰/۰۸۷	۰/۰۸۲	
۰/۳۲	۰/۰۹۰	۰/۰۸۰		۰/۳۵	۰/۰۹۷	۰/۰۸۳		۰/۳۸	۰/۱۰۵۰	۰/۰۷۲	میانگین خطا
۰/۰۰۹	۰/۰۰۶	۰/۰۰۶		۰/۰۳	۰/۰۱۲	۰/۰۰۰		۰/۰۶	۰/۰۲۲	۰/۰۰۹	انحراف معیار

جدول (۲) نتایج حاصل از بازآموزی نه شبکه مورد استفاده در روش این مقاله (ج) با مقادیر ضرایب انتظام یافته شده به طور ثابت در تمام طول بازآموزی.

خطاهای حاصل از بازآموزی			خطاهای حاصل از روش ج			
دستهٔ آزمایشی۲	دستهٔ آزمایشی۱	دستهٔ آموزشی	دستهٔ آزمایشی۲	دستهٔ آزمایشی۱	دستهٔ آموزشی	
۰/۳۹	۰/۱۰۱	۰/۰۹۳	۰/۳۲	۰/۰۹۰	۰/۰۹۳	
۰/۴۲	۰/۱۱۲	۰/۰۸۹	۰/۳۳	۰/۱۰۴	۰/۰۷۲	
۰/۴۴	۰/۱۰۴	۰/۰۹۱	۰/۳۱	۰/۰۹۶	۰/۰۷۳	
۰/۳۶	۰/۰۹۱	۰/۰۸۶	۰/۳۴	۰/۰۹۳	۰/۰۸۵	
۰/۳۳	۰/۰۸۸	۰/۰۷۹	۰/۳۲	۰/۰۸۶	۰/۰۸۱	
۰/۴۰	۰/۱۰۲	۰/۰۷۱	۰/۳۱	۰/۰۹۰	۰/۰۷۶	
۰/۳۶	۰/۰۸۶	۰/۰۸۳	۰/۳۲	۰/۰۸۳	۰/۰۷۹	
۰/۳۵	۰/۰۹۱	۰/۰۸۴	۰/۳۳	۰/۰۸۸	۰/۰۷۸	
۰/۳۲	۰/۰۸۲	۰/۰۸۵	۰/۳۳	۰/۰۸۲	۰/۰۸۷	
۰/۳۷	۰/۰۹۵	۰/۰۸۴	۰/۳۲	۰/۰۹۰	۰/۰۸۰	میانگین خطا
۰/۰۴۰	۰/۰۱	۰/۰۰۶	۰/۰۰۹	۰/۰۰۶	۰/۰۰۶	انحراف معیار

جدول (۳) مقایسه نتایج بدست آمده از روش ارائه شده با نتایج گزارش شده توسط روش‌های گوناگون مطرح گردیده در دیگر مقالات.

	دستهٔ آموزشی	دستهٔ آزمایشی ۱	دستهٔ آزمایشی ۲
نتایج مقالهٔ [۲۸]	۰/۰۸۲	۰/۰۸۶	۰/۳۵
نتایج مقالهٔ [۲۹]	۰/۰۹۰±۰/۰۰۱	۰/۰۸۲±۰/۰۰۷	۰/۳۵±۰/۰۵
نتایج مقالهٔ [۳۰]	۰/۰۸۲±۰/۰۰۱	۰/۰۸۲±۰/۰۰۲	۰/۳۵۷±۰/۰۱۳
نتایج این مقاله	۰/۰۸۰±۰/۰۰۶	۰/۰۹۰±۰/۰۰۶	۰/۳۲±۰/۰۰۹

مراجع

[1] Cybenco G., "Approximation by superpositions of a sigmoidal function", Math. Control Signals Systems, 2, pp. 183-192, 1989.

[2] Funahashi K., "On the approximation realization of continuous mappings by neural networks", Neural Networks, vol. 2, pp. 183-192, 1989.

[3] Hornik K., Stinchcombe M., White H., "Multilayer feedforward networks are universal approximators", Neural Networks, vol. 2, pp. 359-366, 1989.

[4] Bishop C. M., "Neural Networks for pattern recognition", Oxford, UK: Oxford University Press, 1995.

[5] Reed R., "Pruning algorithms - A survey", IEEE Transaction on Neural Networks, ITNN, vol. 4, No. 5, pp. 740-747, 1993.

[6] Levin A. U., Leen T. K., Moody J. E., "Fast pruning using principal components", NIPS'6, Cowan, Tesauro, and Alspector (editors), Morgan Kaufmann Publishers, San Mateo, 1994. URL: http://www. cse.ogi.edu/moody/pubs.html.

[7] Thodberg H. H., "Improving generalization of neural networks through pruning", International Journal of Neural Systems 1, pp. 317-326, 1991.

[8] Sarle W. S., "Stopped training and other remedies for overfitting", Proceedings of the 27th Symposium on the interface of Computing Science and Statistics, pp. 325-360, 1995. URL: ftp://ftp. sas. com/pub/neural/inter 95. ps.Z

[9] Poggio T., Girosi F., "Networks for approximation and learning", Proc. IEEE, 78 (9), pp. 1481-1497, 1990.

[10] Larsen J., Svarer C., Andersen L. N., Hansen L. K., "Adaptive regularization in neural network modeling", to appear in Orr G. B., Muller K., Caruana R. (editors), The Book of Trics, Germany: Spring-Verlag, 1997. URL: http://eivind. imm. dtu. dk/staff/jlarsen/CV. html.

[11] Rissanen J., "Modeling by shortest data discription", Automatica 14, pp. 465-471, 1978.

[12] Rissanen J., " Stochastic comlexity and modeling", Annals of statistics 14, pp. 1080-1100, 1986.

[13] Hansen S. J., Pratt L. Y., "Comparing biases for minimal network construction with backpropagation", In. D. S. Touretzky, (editor), NIPS'1, pp. 177-185, Morgan Kaufmann, 1989.

[14] Hinton G. E. "Connectionist learning procedures", Technical Report, GMU-CS-87-115 (version 2), Carnegie-Melon University, 1987.

[15] Krogh A., Hertz J. E., "A simple weight decay can improve generalization", NIPS'4, Moody J.E., Hanson S. J., Lippmann R. P. (editors), Morgan Kaufmann publishers, San Mateo CA, pp. 950-957, 1995.

[16] Weigend A. S., Huberman B. A., Rumelhart D. E., "Predicting sunspot and exchange rates with connectionist networks", In. Casdagli M. and Eubank S. (editors), Nonlinear Modeling and Forecasting, SFI studies in the science of complexity, vol. 12, Addison-wesely, 1991.

[17] Weigend A. S., Rumelhutr D. E., Huberman B.

A., "Back-propagation, weight elimination and time series prediction", In Proceedings of the 1990 Connectionist Models Summer School, pp. 65-80, Morgan Kaufmann, 1990.

[18] Weigend A. S., Rumelhart D. E., Huberman B. A., "Generalization by weight elimination applied to currency exchange rate prediction", IJCNN, pp. 1837-1841, Seattle, 1991.

[19] Hansen S. J. , Pratt . L. Y., "Some comparisons of constraints for minimal network construction with back-propagation", In D. Touretzky (editor), NIPS, San Mateo, CA: Morgan Kaufmann, 1989.

[20] Bos S., "Optimal weight decay in a perceptron", In International Conference on Artificial Neural Networks (ICANN, 96-Bochum), pp. 551-556, 1996.

URL: http://www. bip. riken. go. jp/irl/boes/ sbopub. html

[21] Bos S., Chng E., "Using Weight decay to optimize the generalization ability of a perceptron", ICNN, pp. 241-246, 1996.

[22] Hintz-Madsen M., Pedersen M. W., Hansen L. K., Larsen J., "Design and evaluation of neural classifiers", Proceedings of the IEEE Workshop on Neural Networks for Signal Processing VI, Piscataway, New Jersey : IEEE, 1996. URL: http://www.ei.dtu.dk/staff/hintz/www/NN/nnpublic. html.

[23] Larsen J., Hansen L. K., Svarer C., Ohlsson M., "Design and regularization of neural networks: The optimal use of validation set", In Usui S., Tohkura Y., Katagiri S., Wilson E. (editors). Proceedings of the IEEE workshop on Neural Networks for Signal Processing VI, Piscataway, New Jersey:IEEE, pp. 62-71, 1996.

URL: http://eivind. imm. dtu. dk/staff/jlarsen/CV.html.

[24] Larsen J., Hansen L. K., Svarer C., "Regularization of neural networks", In Hansen P. C. (editor), Proceedings of the 4th Interdisciplinary Workshop, Technical University of Denmarks, pp. 59-66, 1996.

URL: http://eivind. imm. dtu. dk/staff/jlarsen/CV. html.

[25] Andersen L. N., Larsen J., Hansen L. K., Hintz Madsen M., "Adaptive regularization of neural classifiers", In Principe J., Giles L., Morgan N., Wilson E. (editors), Proceedings of the IEEE Workshop on Neural Networks for Signal Processing VII, Piscataway, New Jersey: IEEE, pp. 24-33, 1997. URL: http:// eivind. imm. dtu. dk/staff/jlarsen/CV. html

[26] Becker S., Le Cun Y., "Improving the convergence of backpropagation learning with second order methods", In Touretzky D., Hinton G. E. , Sejnowski T. J. (editors), Proceedings of the 1988 Connectionist Models Summer School, pp. 29-37, San Mateo, CA: Morgan Kaufmann, 1989.

[27] Larsen J., "Design of neural network filters", Ph.D. Thesis, Electronics Institue, Technical University of Denmark, (435 pages), 1993. Second Edition 1996. URL: http ://eivind. imm. dtu. dk/staff/jlarsen/CV. html.

[28] Weigend A. S., Huberman B. A., Rumelhart D. E., "Predicting the future: a connectionist approach", International Journal of Neural Systems, 1(3), pp. 193-210, 1990.

[29] Svarer C., Hansen L. K., Larsen J., "On design and evaluation of tapped-delay neural network architectures", In Berenji H. R. (editor), IEEE International Conference on Neural Networks, ICNN, pp. 46-51, Piscataway, NJ, 1993.

[30] Goutte C., "On the use of a pruning prior for neural networks", Neural Network for Signal Processing VI, pp. 52-61, 1996. URL: http:// eivind . imm. dtu. dk/staff / goutte / PUBLIS/ nnsp 96. html.