جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

دانشگاه صنعتی امیر کبیر

(پلی تکنیک تهران)

بهبود سرعت یادگیری شبکه های عصبی چند لایه با الگوریتم پس انتشار خطا

گردآورنده:

مهدیه سادات سعدآبادی   ۸۳۱۲۳۱۳۹

شهریور ماه ۸۴

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

www.kandoocn.com

www.kandoocn.com

www.kandoocn.com

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

www.kandoocn.com

بهبود سرعت یادگیری شبکه های عصبی چند لایه با الگوریتم پس انتشار خطا

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

شبکه های عصبی چند لایه پیش خور[1] به طور وسیعی د ر زمینه های متنوعی از قبیل طبقه بندی الگوها، پردازش تصاویر، تقریب توابع و ... مورد استفاده قرار گرفته است.

الگوریتم یادگیری پس انتشار خطا[2]، یکی از رایج ترین الگوریتم ها جهت آموزش شبکه های عصبی چند لایه پیش خور می باشد. این الگوریتم، تقریبی از الگوریتم بیشترین تنزل[3] می باشد و در چارچوب یادگیری عملکردی [4] قرار می گیرد.

عمومیت یافتن الگوریتم BP، بخاطر سادگی و کاربردهای موفقیت آمیزش در حل مسائل فنی- مهندسی می باشد.

علیرغم، موفقیت های کلی الگوریتم BP در یادگیری شبکه های عصبی چند لایه پیش خور هنوز، چندین مشکل اصلی وجود دارد:

- الگوریتم پس انتشار خطا، ممکن است به نقاط مینیمم محلی در فضای پارامتر، همگرا شود. بنابراین زمانی که الگوریتم BP همگرا می شود، نمی توان مطمئن شد که به یک جواب بهینه رسیده باشیم.

- سرعت همگرایی الگوریتم BP، خیلی آهسته است.

از این گذشته، همگرایی الگوریتم BP، به انتخاب مقادیر اولیه وزنهای شبکه، بردارهای بایاس و پارامترها موجود در الگوریتم، مانند نرخ یادگیری، وابسته است.

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

انجام شده نیز نشان می دهد، الگوریتم های پیشنهادی نسبت به الگوریتم استاندارد BP، از سرعت همگرایی بالاتری برخوردار هستند.

خلاصه ای از الگوریتم BP

از قانون یادگیری پس انتشار خطا (BP)، برای آموزش شبکه های عصبی چند لایه پیش خور که عموماً شبکه های چند لایه پرسپترون[5] (MLP) هم نامیده می شود، استفاده می شود، استفاده می کنند. به عبارتی توپولوژی شبکه های MLP، با قانون یادگیری پس انتشار خطا تکمیل می شود. این قانون تقریبی از الگوریتم بیشترین نزول (S.D) است و در چارچوب یادگیری عملکردی قرار می گیرد.

بطور خلاصه، فرایند پس انتشار خطا از دو مسیر اصلی تشکیل می شود. مسیر رفت[6] و مسیر برگشت[7].

در مسیر رفت، یک الگوی آموزشی به شبکه اعمال می شود و تأثیرات آن از طریق لایه های میانی به لایه خروجی انتشار می یابد تا اینکه

---

1. Multi-Layer Feedforward Neural Networks
2. Back-Propagation Algorithm
3. Steepest Descent (S.D)
4. Performance Learning
5. Multi Layer Perceptron
6. Forward Path
7. Backward Path

نهایتاً خروجی واقعی شبکه MLP، به دست می آید. در این مسیر، پارامترهای شبکه (ماتریس های وزن و بردارهای بایاس)، ثابت و بدون تغییر در نظر گرفته می شوند.

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

بر اساس قانون یادگیری اصلاح خطا[1] انجام می گیرد. سیگنال خطا، رد لایه خروجی شبکه تشکیل می گردد. بردار خطا برابر با اختلاف بین پاسخ مطلوب و پاسخ واقعی شبکه می باشد. مقدار خطا، پس از محاسبه، در مسیر برگشت از لایه خروجی و از طریق لایه های شبکه به سمت پاسخ مطلوب حرکت کند.

در شبکه های MLP، هر نرون دارای یک تابع تحریک غیر خطی است که از ویژگی مشتق پذیری برخوردار است. در این حالت، ارتباط بین پارامترهای شبکه و سیگنال خطا، کاملاً پیچیده و و غیر خطی می باشد، بنابراین مشتقات جزئی نسبت به پارامترهای شبکه به راحتی قابل محاسبه نیستند. جهت محاسبه مشتقات از قانون زنجیره ای[2] معمول در جبر استفاده می شود.

فرمول بندی الگوریتم BP

الگوریتم یادگیری BP، بر اساس الگوریتم تقریبی SD است. تنظیم پارامترهای شبکه، مطابق با سیگنالهای خطا که بر اساس ارائه هر الگو به شبکه محاسبه می شود، صورت می گیرد.

الگوریتم بیشترین تنزل با معادلات زیر توصیف می شود:

$$W^{L} ji(K+1) = Wji^{L}(K) - \alpha \frac{\delta F}{\delta W^{L} ji(k)} \qquad (1)$$

$$b^{L} j(K+1) = bj^{L}(K) - \alpha \frac{\delta F}{\delta bj^{L}(k)} \qquad (2)$$

به طوری $W^{L}ji$ و $b^{L}j$، پارامترهای نرون j ام در لایه iام است. α، نرخ یادگیری[2] و F، میانگین مربعات خطا می باشد.

www.kandoocn.com

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

$\delta W^{L} ji(k)$

$$\frac{\delta F(k)}{\delta b^{L} j(k)} = S^{L} j(k) \qquad (4)$$

$$S^{L} j(k) = \frac{\delta F(K)}{\delta nj^{L}(K)} = \left[\sum_{m=1}^{S^{L+1}} S_{m}^{\ l+1}(k) W_{mj}^{l+1}(k)\right] f^{L}(N_{j}^{L}(k) \qquad (5)$$

به طوریکه $S^{L}_{j}(k)$، حساسیت رفتار شبکه در لایه L ام است.

---

1. Error-Correctting Learning Rule
2. Chain Rule
3. Learning Rate

معایب الگوریتم استاندارد پس انتشار خطا[1] (SBP)

الگوریتم BP، با فراهم آوردن روشی از نظر محاسباتی کارا، رنسانسی در شبکه های عصبی ایجاد نموده زیرا شبکه های MLP، با قانون یادگیری BP، بیشترین کاربرد را در حل مسائل فنی- مهندسی دارند. با وجود، موفقیت های کلی این الگوریتم در یادگیری شبکه های عصبی چند لایه پیش خود، هنوز مشکلات اساسی نیز وجود دارد:

- اولاً سرعت همگرایی الگوریتم BP آهسته است.

همانطور که می دانیم، تغییرات ایجاد شده در پارامترهای شبکه (ماتریس های وزن و بردارهای بایاس)، پس از هر مرحله تکرار الگوریتم BP، به اندازه $\alpha\nabla F(x(k))$، است، به طوریکه F، شاخص اجرایی، x پارامترهای شبکه و α، طول قدم یادگیری است.

از این، هر قدر طول قدم یادگیری، α، کوچکتر انتخاب گردد، تغییرات ایجاد شده در پارامترهای شبکه، پس از هر مرحله تکرار الگوریتم BP، کوچکتر خواهد بود، که این خود منجر به هموار گشتن مسیر

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

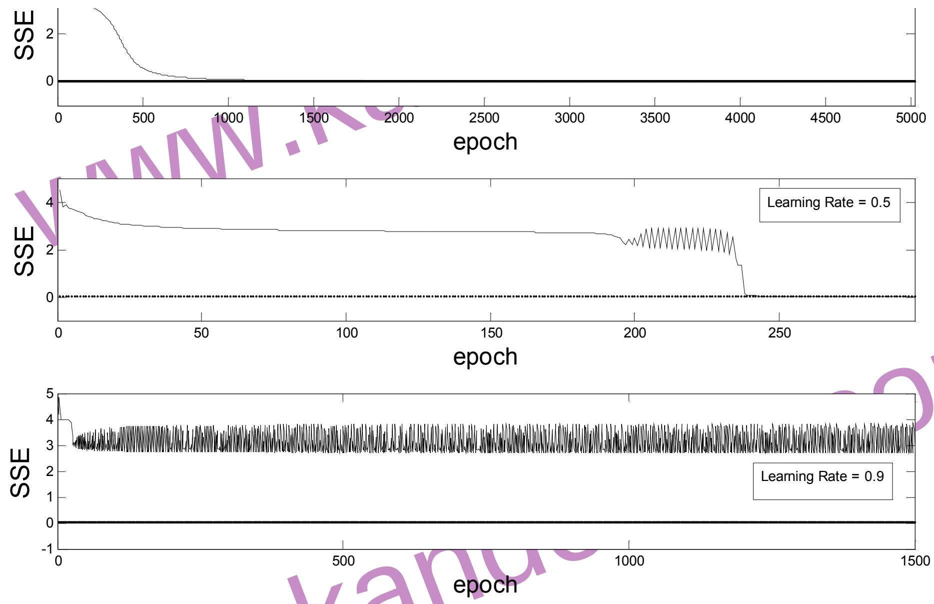
الگوریتم BP می گردد. بر عکس با افزایش طول قدم α، اگرچه نرخ یادگیری و سرعت یادگیری الگوریتم BP افزایش می یابد، لیکن تغییرات فاحشی در پارامترهای شکه از هر تکراربه تکرار بعد ایجاد می گردد، که گاهی اوقات موجب ناپایداری و نوسانی شدن شبکه می شود که به اصطلاح می گویند پارامترهای شبکه واگرا شده اند:

در شکل زیر، منحنی یادگیری شبکه برای جدا سازیالگوها در مسأله XOR، به ازای مقادیر مختلف نرخ یادگیری، نشان داده شده است. به ازای مقادیر کوچک، α، شبکه کند اما هموار، یاد نمی گیرد الگوهای XOR را از هم جدا نماید، ددر صورتی که به ازای ۰/۹= α شبکه واگرا می شود.

1. Standard Back-Propagation Algorithm

www.kandoocn.com

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید



شکل (۱). منحنی یادگیری شبکه برای نرخ های یادگیری مختلف در مسأله XOR

– ثانیاً احتمالاً به دام افتادن شبکه در نقاط مینیمم محلی وجود دارد.

در شبکه های MLP، میانگین مجوز خطا، در حالت کلی خیلی پیچیده است و از تعداد زیادی نقطه اکسترمم در فضای پارامترهای شبکه برخوردار می باشد. بنابراین الگوریتم پس انتشار خطا با شروع از روی یک سری شرایط اولیه پارامترهای شبکه، به نقطه مینیمم سراسری و با شروع از یک مجموعه شرایط اولیه دیگر به تقاط مینیمم محلی در فضای پارامترها همگرا می گردد، بنابراین زمانی که الگوریتم BP همگرا می شود، نمی توان مطمئن شد که به یک جواب بهینه رسیده باشیم.

– ثالثاً: همگرایی الگوریتم BP، به یقین مقادیر اولیه پارامترهای شبکه عصبی MLP وابسته است، بطوری که یک انتخاب خوب می تواند کمک بزرگی در همگرایی سریعتر الگوریتم BP فراهم آورد.

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

در فضای برداری پارامترهای شبکه می گردد که این خود منجر به این می شود که شبکه خیلی زودتر از معمول به موضعی بیفتد که منحنی یادگیری شبکه برای تعداد بزرگی از دفعات تکرار، تغییر نکند.

به عنوان مثال، فرض می کنیم مقدار اولیه پارامترهای شبکه خیلی بزرگ باشند، در حالی که می دانیم توابع تبدیل نرونها مخصوصاً در لایه های میانی از نوع زیگموئید هستند. در این حالت برای نرون i ام، اندازه ورودی تابع تبدیل (ni) خیلی بزرگ می باشد و خروجی نرون (ai) به مقدار ±۱ میل می کند. لذا مشتق بردار خروجی شبکه، a، خیلی کوچک می باشد. فرض کنیم که باید مقدار واقعی ai، ۱ باشد یعنی $ti = 1$، لیکن به خاطر انتخاب بر مقادیر اولیه، $ai = -1$ گردد. در این حالت خطای حداکثر را داریم در حالی که چون $ai \approx 0$ می باشد تغییرات ناچیزی در پارامترهای متناظر با نرون i ام داریم. این چیزی است که بیانگر رسیدن زودتر از معمول نرونها به حد اشباع خود می باشند، جایی که پاسخ واقعی با پاسخ شبکه کاملاً فرق دارد و زمان زیادی طول خواهد کشید که نرون از این حالت خارج شود. از این رو با پیشرفت پروسه یادگیری، پارامترهای منتسب به نرورنهایی که به مرز اشباع نرسیده اند، سریعتر تنظیم می شوند، چرا که سیگنال خطار گرادیانهای محلی از مقدار از اندازه بزرگتری برخوردار می باشند. این عمل منجر به کاهش در مجموع مربعات خطای لحظه ای می گردد و اگر در این مرحله، نرونهای به حد اشباع رسیده تغییری در وضعیت تحریکشان رخ ندهد، شبکه برای مدتی طولانی از یک شکل هموار منحنی خطا برخوردار خواهدبود.

بهبود الگوریتم استاندارد پس انتشار خطا (SBP)

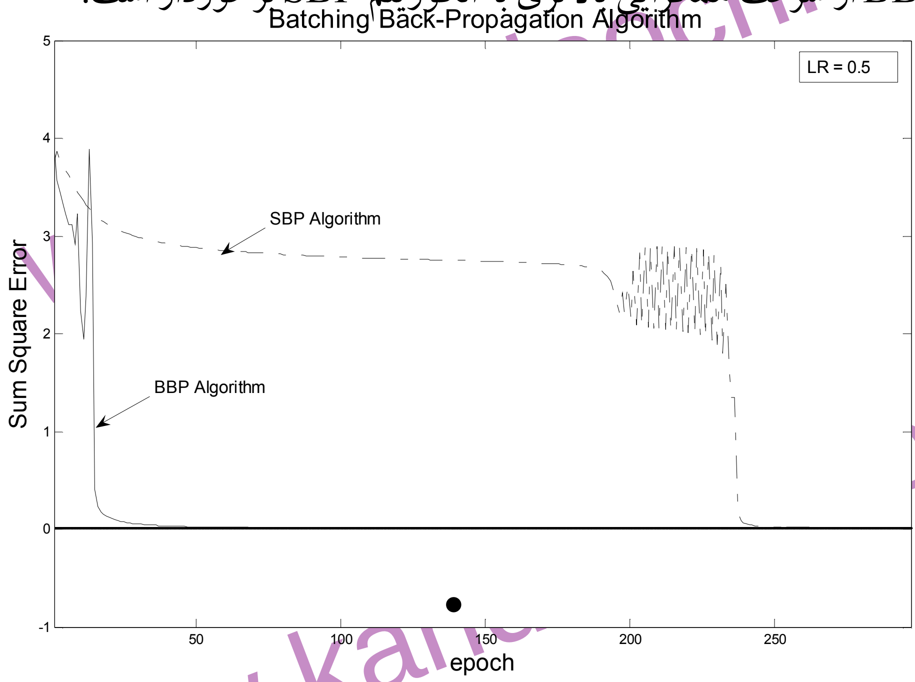
- الگوریتم BP از نوع دسته ای[1] (BBP)

الگوریتم استاندارد BP، بر اساس فرم الگو به الگو است، بدین ترتیب که پارامترهای شبکه پس از ارائه هریک از الگوهای یادگیری که عموماً بطور تصادفی انتخاب می شوند، تنظیم می گردند، اما در

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

دسته ای موجب می شود که گرادیانهای محلی به گرادیان محلی واقعی نزدیکتر باشند و نهایتاً الگوریتم BP به الگوریتم بیشترین نزول نزدیکتر گردد که این خود موجب می شود همگرایی الگوریتم BP افزایش یابد.

در شکل زیر مسئله XOR با متد الگوریتم BP به فرم دسته ای پیاده شده است. به راحتی می توان دید که الگوریتم BBP از سرعت همگرایی بالاتری به الگوریتم SBP برخوردار است.



شکل (۲). رفتار شبکه با الگوریتم BBP در مسأله XOR (ـــ)
رفتار شبکه با الگوریتم SBP (ـ٠ـ)

1. Batching Back-Propagation Algorithm

– روش ممنتم [1] برای الگوریتم BP (MBP)

www.kandoocn.com

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

الگوریتم SD است، بسیار کند می گردد. و اگر α، بزرگتر انتخاب شود، شبکه نوسانی خواهد بود.

یک راه ساده و مؤثر که عموماً جهت افزایش و بهبود نرخ یادگیری، استفاده می شود- جایی که خطر ناپایداری و نوسانی شدن شبکه جلوگیری می گردد- افزودن یک جمله ممنتم در الگوریتم تقریبی SD می باشد، یعنی به هر پارامتر از شبکه MLP، یک مقدار اینرسی یا اندازه حرکت اضافه می شود تا اینکه پارامتر مورد نظر در مسیری تمایل به تغییر داشته باشد که کاهش تابع انرژی احساس شود.

الگوریتم یادگیری MBP با معادلات زیر قابل توصیف است:

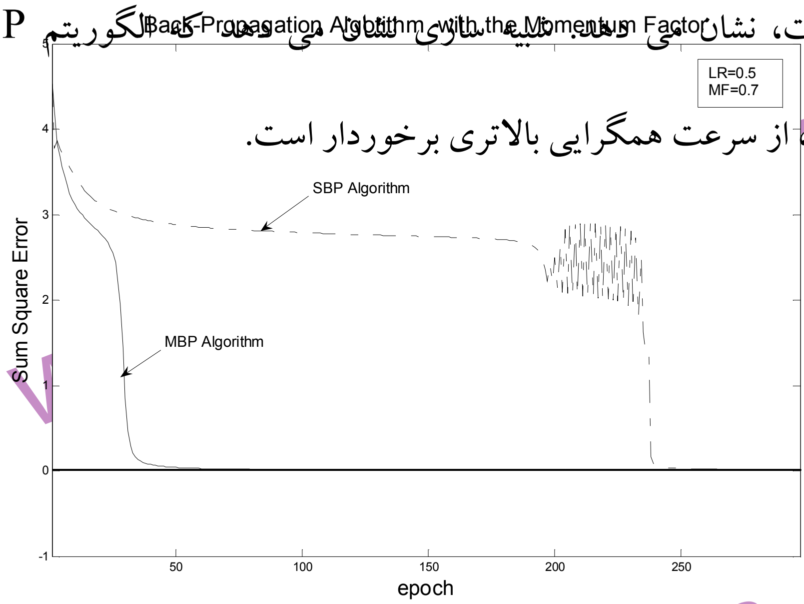
$$(6)\ \Delta W^{L}(k) = \delta\ \ \Delta w^{L}(k-1) - \beta\delta^{L}(a^{L-1})^{T}$$

$$(7)\ \Delta \underline{b}^{L}(k) = \delta\ \ \Delta \underline{b}^{L}(k-1) - \beta\delta^{L}$$

جایی که $\delta \in (0و1)$، ترم ممنتم را نشان می دهد و عموماً با نرخ یادیگیری δ,α به صورت زیر رابطه دارد:

$$(8)\ \beta = \alpha(1-\delta)$$

معادلات فوق، ترم های اصلاحی پارامترهای شبکه را از فیلتر پایین گذر عبور می دهند واین یعنی تغییرات با فرکانس بالا(نوسانات شدید) فیلترمی شوند. شکل (۳)، مسأله XOR را که به وسیله الگوریتم MBP، پیاده شده است، نشان می دهد. شبیه سازی نشان می دهد که الگوریتم MBP نسبت به الگوریتم استاندارد BP، از سرعت همگرایی بالاتری برخوردار است.



www.kandoocn.com

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

شکل (۳): رفتار شبکه با الگوریتم MBP درمسأله XOR (ـــ)
رفتار شبکه با الگوریتم SBP (ـــ.)

1. Momentum

- نرخ یادگیری متغیر [1] (VLR)

درالگوریتم BP استاندارد، نرخ یادگیری درطول فرآیند یادگیری ثابت نگه داشته می شود. عملکرد الگوریتم به انتخاب مناسب نرخ یادگیری خیلی حساس می باشد. اگر نرخ یادگیری خیلی بزرگ انتخاب شود ممکن است الگوریتم نوسان کرده وناپایدار شود. اگر نرخ یادگیری خیلی کوچک باشد زمان زیادی طول خواهد کشید تا الگوریتم همگرا شود. انتخاب نرخ یادگیری اپتیمم قبل از یادگیری، عملی نبوده ودرحقیقت نرخ یادگیری اپتیمم به هنگام پروسه آموزش، همچنان که الگوریتم برروی سطح خطا حرکت می کنددائماً تغییر می کند.

اگر اجازه دهیم نرخ یادگیری بهنگام پروسه آموزش تغییر کند عملکرد الگوریتم BP استاندارد را می توان بهبود بخشید. نرخ یادگیری تطبیقی سعی می کند که نرخ یادگیری را تا آنجایی که ممکن است و سیستم ناپایدار نشده است، افزایش دهد.

نرخ یادگیری تطبیقی نیاز به تغییراتی در الگوریتم BP استاندارد دارد.

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

```
VLR Algorithm
Initialize Nearal Neywork Weights and Biases.
Set Training Parameters.
for    i = 1: N
; break, endε          if  SSE <
          Feed forward Path;
          Backward Path;
          Compute New Weights and Biases;
```

```
           Compute  new-SSE
         if    new-SSE > SSE * max – perf-inc
               α = α * Lr-dec;       % Learning Rate
               δ = 0               ;       % Momentum Factor
         else
               if new-SSE < SSE
               α = α * Lr-inc;
               end
         Compute New Weights and Biases
                 end
          end
```

1. Variable Learning Rate

معمولاً مقادیر 1.04، 0.7 و 1.05 به ترتیب برای ضرائب [1] max –perf –inc، [2] Lr –dec و [3] Lr –inc در نظر گرفته می شود.

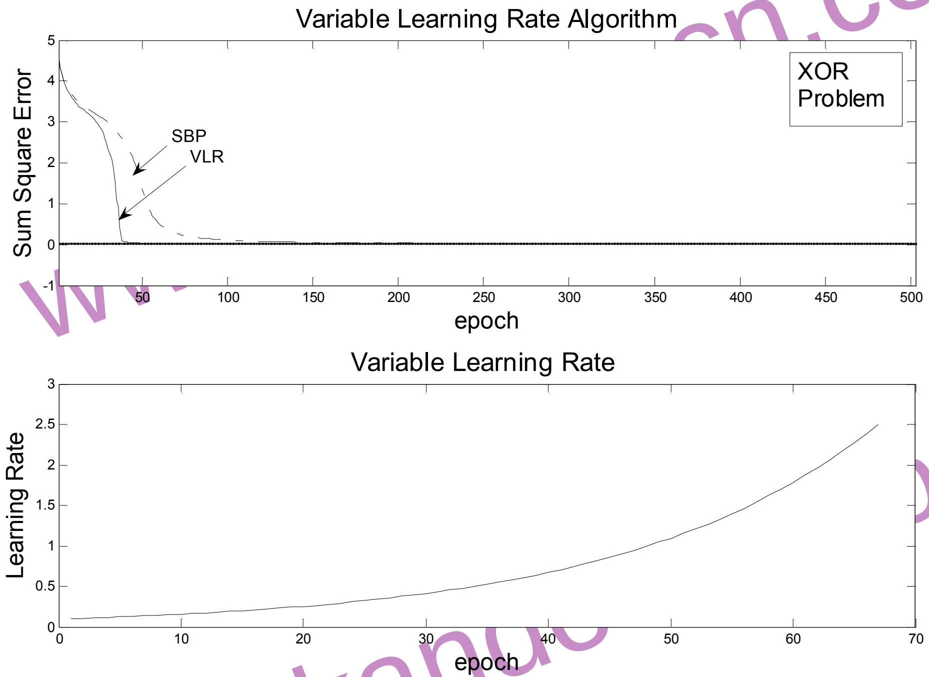
www.kandoocn.com

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

نگردد. بنابرین، یک نرخ یادگیری نزدیک به بهینه بدست می آید.

الگوریتم VLR بر روی مسأله XOR پیاده شده است. شبیه سازی نشان می دهد، که سرعت یادگیری این الگوریتم را از الگوریتم SBP، بیشتر است.

نتایج شبیه سازی در شکل (۴) نشان داده شده است.



شکل (۴). – رفتار شبکه با الگوریتم VLR برای مسأله XOR(ــــ)

رفتار شبکه با الگوریتم SBP(۰ ــــ)

– تغییرات نرخ یادگیری (α) در کل فرآیند یادگیری برای مسأله XOR

در این، خلاصه ای از نتایج مقالات مورد بررسی، جهت بهبود الگوریتم یادگیری پس انتشار خطا، ذکر می گردد:

www.kandoocn.com

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

اگر از نرخ یادگیری کوچک استفاده شود، این مسأله می تواند سبب نرخ پائین همگرایی. به منظور جلوگیری از این پدیده نامطلوب، یک راه حل، استفاده از نرخ یادگیری می باشد. این راه حل بویژه در مواقعی که در نقطه ای از سطح قرار داریم که شیب تندی دارد، نامطلوب بوده و می تواند باعث واگرایی شبکه شود.

---

1. Maximum Performance Increase
2. Learning Rate decrease
3. Learning Rate increase

موقعیکه شکل سطح خطا از سطح خطای درجه دوم خیلی فاصله داشته باشد، در اینصورت سطح خطا، شامل مناطق با شیب تند زیادی خواهد بود. در این صورت الگوریتم BP با نرخ یادگیری ثابت دارای راندمان پائینی می باشد، دلیل این مسأله این است که به منظور جلوگیری از نوسان در مناطقی که سطح خطا دارای شیب زیادی است بایستی نرخ یادگیری کوچک انتخا شود، در نتیجه بردار وزن، موقعی که در مناطق مسطح قرار داریم به دلیل کوچک بودن گرادیان، خیلی کند حرکت خواهد کرد.

بنابراین نیاز به یک الگوریتم یادگیری کارآمد می باشد تا بتواند بطور پویا نرخ یادگیری را تغییر دهد.

– الگوریتم پس انتشار خطای تطبیقی[1] (ABP)

در الگوریتم پس انتشار خطای تطبیقی، نرخ یادگیری، بطور، اتوماتیک و بر اساس خطای آموزش تنظیم می گردد. [۱]

ایده اصلی از الگوریتم پس انتشار خطای تطبیقی آن است که:

$$\frac{\partial F}{\partial W}$$

www.kandoocn.com

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

خطای جدید نسبت به خطای قبلی، افزایش یافته است و تکرار جاری با ارزش نیست.

لذا نرخ یادگیری را بایستی کاهش داد.

- اگر علامتهای امتدادگرادیان $\frac{\delta F}{\delta W}$ در طول دو تکرار متوالی، یکسان باشند، دلیل بر این امر است که نرخ تنزل آهسته است و نرخ یادگیری را بایستی افزایش داد.

ایده فوق را می توان با معادلات زیر، توصیف کرد:

$$(8)\ \Delta\alpha(K) = \varepsilon\lambda\alpha(K-1)$$

به طوری که $0.01 \le \varepsilon \le 0.1$ و λ را به صورت زیر، تعریف می شود:

$$(9)\ \lambda = \mathrm{Sign}\ \left(\frac{\delta F(K)}{\delta W(K)} \cdot \frac{\delta F(K-1)}{\delta W(K-1)}\right)$$

مشخص است که زمانی که $\lambda>0$، نرخ یادگیری افزایش می یابد.

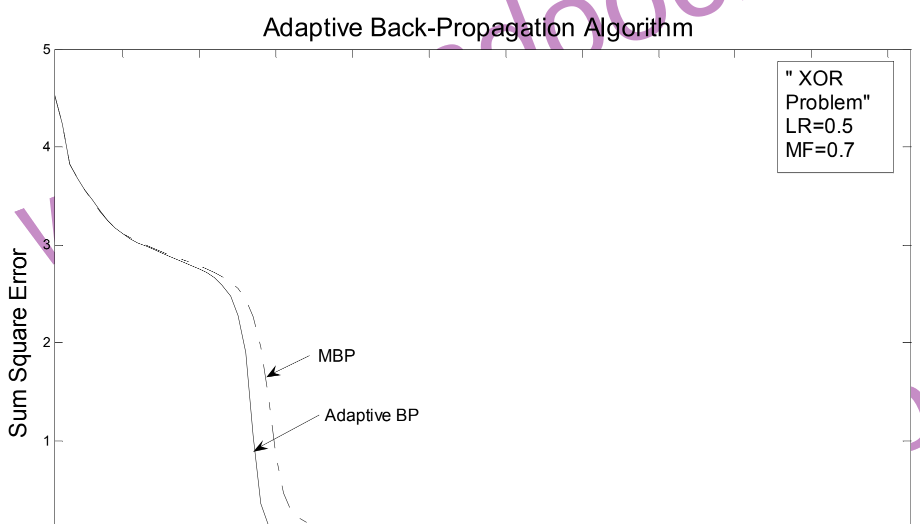
الگوریتم BP تطبیقی با معادلات زیر بیان می شود:

$$(10)\ W(k+1) = W(k) - \alpha(k)\frac{\delta F(k)}{\delta w(k)} + \delta\Delta w(k-1)$$

شکل (۵)، رفتار شبکه را با قانون یادگیری فوق، جهت جداسازی الگئهای XOR، نشان می دهد.

---

1. Adaptive Back-Propagation Algorithm



www.kandoocn.com

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

شکل (۵). منحنی یادگیری الگوریتم BP تطبیقی برای XOR

– الگوریتم پس انتشار خطا با نرخ یادگیری و ضریب ممنتم تطبیقی[1] (BPALM)

در این الگوریتم نرخ یادگیری و ضریب ممنتم در هر سیکل به طور تطبیقی تنظیم می شوند تا به همگرایی الگوریتم BP استاندارد بهبود بخشیده شود[۲].

در ابتدا، فاکتور نسبی $e_r(k)$، به صورت زیر تعریف می شود:

$$e_r(k) = \frac{\Delta F(k)}{f(k)} = \frac{F(k) - F(k-1)}{F(k)} \qquad (11)$$

اصلاح و تنظیمات نرخ یادگیری و ضریب ممنتم به صورت زیر انجام می گیرد:

Case 1: for $e_r(k) < 0$

$$\alpha(k+1) = \alpha(k)\,[1+ue^{-er(k)}]\ ;\quad u\varepsilon\ (0,1) \qquad (12a)$$

$$\delta(k+1) = \delta(k)\,[1+ve^{-er(k)}]\ ;\quad v\varepsilon\ (0,1) \qquad (12b)$$

Case 2: for $e_r(k) \geq 0$

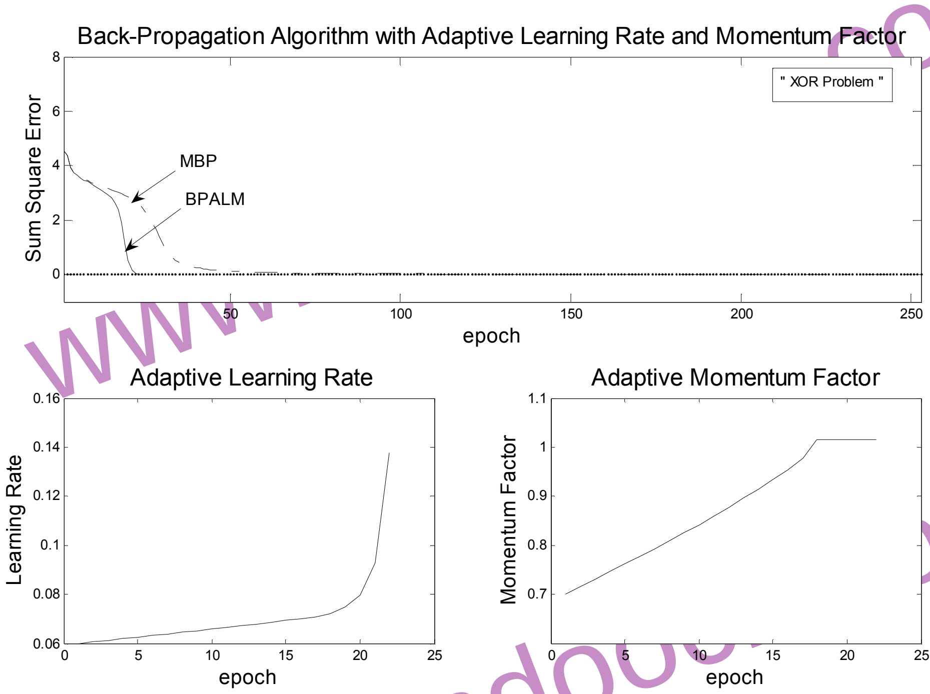
$$\alpha(k+1) = \alpha(k)\,[1-ue^{-er(k)}]\ ;\quad u\varepsilon\ (0,1) \qquad (13a)$$

$$\delta(k+1) = \delta(k)\,[1-ve^{-er(k)}]\ ;\quad v\varepsilon\ (0,1) \qquad (13b)$$

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

1. Back- Propagation with Adaptive Learning rate and Momentum term

شکل زیر، عملکرد شبکه را با قانون یادگیری BPALM، جهت جداسازی الگوهای XOR، نشان می دهد.



شکل(۶). – منحنی یادگیری الگوریتم BPALM در مسأله XOR

- تغییرات نرخ یادگیری

- تغییرات ضریب ممنتم

- تغییرات علامت[1]

بر اساس این الگوریتم، اگر علامت مشتق شاخص اجرایی نسبت به پارامترهای شبکه دردوتکرار متوالی تغییر نکند، نرخ یادگیری افزایش می یابد و در غیر اینصورت، نرخ یادگیری کاهش می یابد. [6]

www.kandoocn.com

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

$$\alpha_{ji}(k) = \alpha_{ji}(k-1).u \quad \text{if} \quad (14\text{ a}) \quad \frac{\delta F(k)}{\delta Wji(k)} \cdot \frac{\delta F(k-1)}{\delta Wji(k-1)} \geq 0$$

$$\alpha_{ji}(k) = \alpha_{ji}(k-1).u \quad \text{if} \quad (14\text{ b}) \quad \frac{\delta F(k)}{\delta Wji(k)} \cdot \frac{\delta F(k-1)}{\delta Wji(k-1)} \prec 0$$

آبگاه، اصلاح وزن بر اساس معادله زیر، انجام می گیرد:

$$(15)\ \Delta W_{ji}(k) = -\alpha_{ji}(k)\frac{\delta F(K)}{\delta Wji(k)} + \delta \Delta W_{ji}(k-1)$$

معموماً مقادیر 1.1-1.3 برای u و 0.7-0.9 برای d، به کار برده می شود.

---

1. Sign Changes

بر اساس این الگوریتم، اگر علامت مشتق شاخص اجرایی، نسبت به پارامترهای شبکه در دوتکرارمتوالی تغییر نکند و نرخ یادگیری از مقدار حداکثری، کمتر باشد، آنگاه افزودن یک مقدار ثابت به نرخ یادگیری، نرخ یادگیری افزیش می یابد. اگر علامت مشتق شاخص اجرایی نسبت به پارامترهای شبکه دردوتکرارمتوالی تغییر کند، نرخ یادگیری با ضرب شدن در مقدار کوچکتر، کاهش می یلبد و در غیر اینصورت تغییری در نرخ یادگیری نخواهیم داشت. [6]

عبارت فوق را می توان با روابط زیر توصیف کرد:

$$\alpha_{ji}(k) = \alpha_{ji}(k-1)+u \quad (16\text{ a}) \quad \frac{\delta F(k)}{\delta Wji(k)} \cdot \frac{\delta F(k-1)}{\delta Wji(k-1)} \geq 0 ,\ (\alpha_{ji}(k-1) \leq \alpha \max$$

if

$$\alpha_{ji}(k) = \quad (16\text{ b}) \quad \frac{\delta F(k)}{\delta Wji(k)} \cdot \frac{\delta F(k-1)}{\delta Wji(k-1)} \leq 0$$

$$\alpha_{ji}(k-1).d \quad \text{if}$$

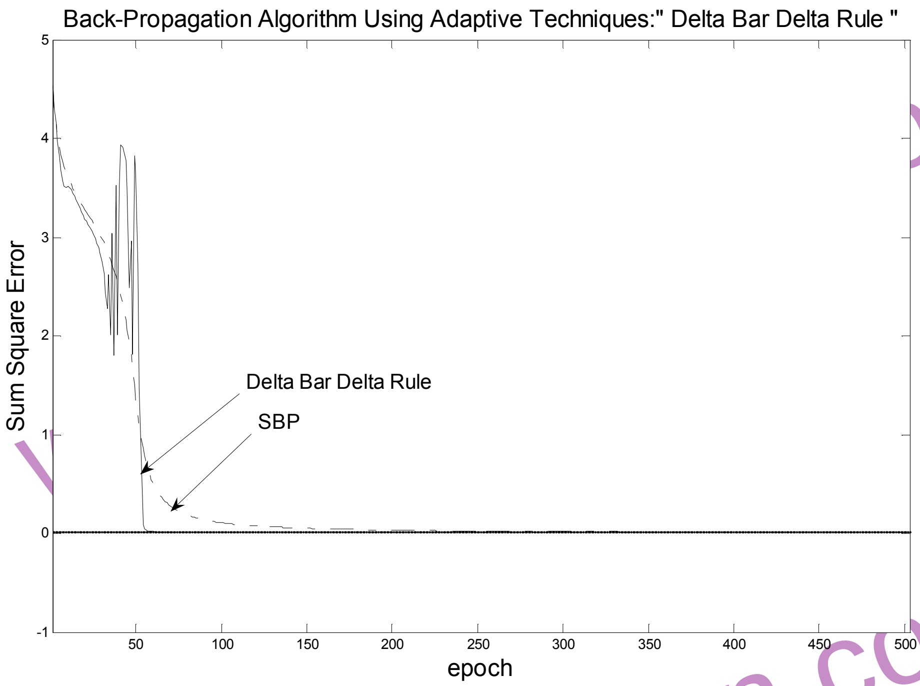
www.kandoocn.com

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

$\delta W_{ji}(k) \quad \delta W_{ji}(k-1)$

$\alpha_{ji}(k-1)$ if

شکل زیر، منحنی یادگیری الگوریتم فوق را برای جداسازی الگوهای XOR، نشان می دهد.



شکل (۷). منحنی یادگیری الگوریتم Delta Bar Delta Rule در مسأله XOR

- الگوریتم یادگیری Super SAB

این الگوریتم، مانند الگوریتم قبل می باشد با این تفاوت که اگر علامت تغییرات مشتق شاخص اجرایی نسبت به پارامترهای شبکه دردوتکرار متوالی تغییر نکند و نرخ یادگیری از مقدار حداکثری کمتر باشد، نرخ یادگیری با ضرب شدن در یک پارامتر ثابتی، افزایش می یابد [6].

www.kandoocn.com

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

$\frac{\delta F(k)}{\delta Wji(k)} \cdot \frac{\delta F(k-1)}{\delta Wji(k-1)}$

if

$$\alpha_{ji}(k) = \qquad (17\text{ b}) \quad \frac{\delta F(k)}{\delta Wji(k)} \cdot \frac{\delta F(k-1)}{\delta Wji(k-1)} \leq 0$$

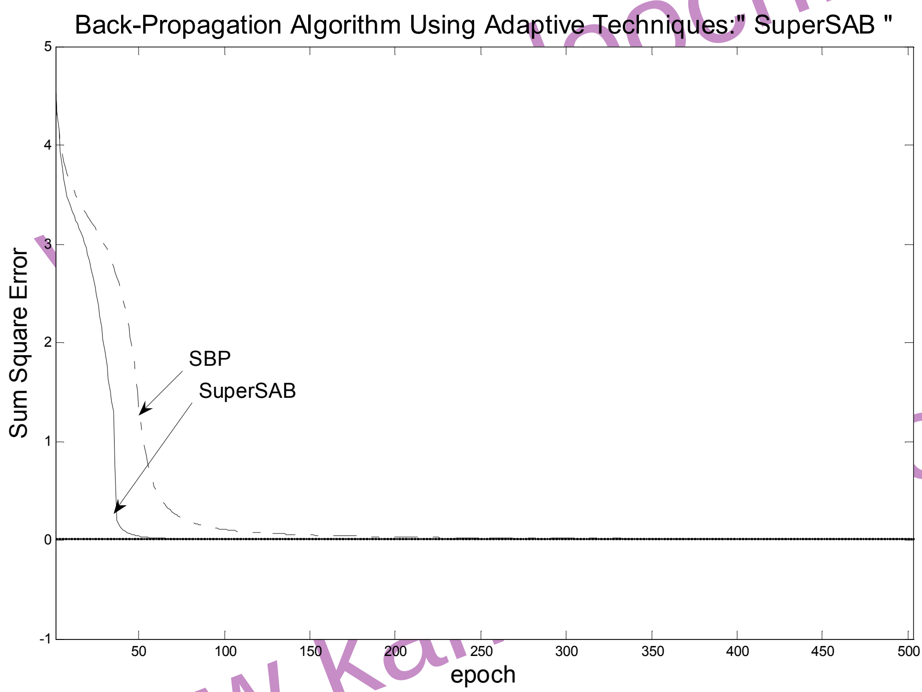
$\alpha_{ji}(k-1).d$ if

$$\alpha_{ji}(k) = \qquad (17\text{ c}) \quad \frac{\delta F(k)}{\delta Wji(k)} \cdot \frac{\delta F(k-1)}{\delta Wji(k-1)} \succ 0$$

$\alpha_{ji}(k-1)$ if

مقادیر پیشنهادی بر u، 1.05 و d، 0.5 می باشد. α_{max}، عددی بین 1,0 است.

از الگوریتم مطرح شده، برای جداسازی الگوها در مسأله XOR، استفاده شده است. در شکل (۸)، منحنی یادگیری این الگوریتم را نشان می دهد.



www.kandoocn.com

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

شکل (۸). منحنی یادگیری الگوریتم Super SAB برای مسأله XOR

www.kandoocn.com

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

در این الگوریتم، ترم جدیدی به نام ضریب تناسبی [1] (PF)، به الگوریتم استاندارد پس انتشار خطا اضافه شده است. [3]

افزایش سرعت یادگیری، همراه با حفظ سادگی اگوریتم BP، هدف اصی این مقاله است.

الگوریتم مطرح شده را می توان با معاله زیر توصیف کرد:

$$(18)\ \Delta W(k) = -\alpha \nabla F(W(k)) + \delta \Delta W(k-1) + ce(W(k))$$

بطوری که C، ضریب تناسبی (PF) می باشد.

e(W(k))، اختلاف بین خروجی مطلوب و خروجی واقعی در تکرار k ام است.

الگوریتم فوق، درای سه ترم است:

۱- ترمی که متناسب با مشتق F(W(k)) است.

۲- ترمی که متناسب با مقادیر قبلی تغییرات وزن هاست.

۳- ترمی که متناسب با e(W(k)) است.

بنابراین، مشاهده می شود که این سه ترم، همانند ضرایب کنترل کننده PID، که به طور متداول در کاربردهای کنترلی استفاده می شوند، عمل می کنند.

آنالیز همگرایی

در این قسمت، الگوریتم BP با سه ترم، آنالیزو تحلیل می شود و نشان داده می شود که نقاط مینیمم محلی تابع حداقل مربعات خطا، تنها نقاط پایدار مجانبیمحلی الگوریتم هستند.

معادله (18) را می توان به صورت زیر، بازنویسی کرد:

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

فرض کنید $x_1(k) = W(k)$, $x_2(k) = W(k) - W(k-1)$، آنگاه یک تحقق متغیر حالت معادله (19)، چنین است:

$$(20a)\quad x_1(k+1) = x_2(k) - \alpha\nabla F(x_1(k)) + \delta x_2(k) + Ce(x_1(k))$$

$$(20b)\quad x_2(k+1) = -\alpha\nabla F(x_1(k)) + \delta x_2(k) + Ce(x_1(k))$$

لم (۱):

$m = (m_1,m_2)$، یک نقطه تعادل از سیستم یا معادلات (20a) , (20b) است، اگر

$$m_2 = 0, \alpha\nabla F(x_1(k)) = Ce(x_1(k))$$

---

1. Proproortional Factor

اثبات:

اگر $m = (m_1,m_2)$، یک نقطه تعادل است، آنگاه

$x_1(k) = m_1$, $x_2(k) = m_2$

و

$x_1(k+1) - x_1(k) = 0$

$x_2(k+1) - x_2(k) = 0$

با جایگذاری در معادلات (20a) , (20b)، خواهیم داشت:

$$(21a)\quad (1-\delta)x_2(k) = -\alpha\nabla F(x_1(k)) + Ce(x_1(k))$$

$$(21b)\quad -\delta x_2(k) = -\alpha\nabla F(x_1(k)) + Ce(x_1(k))$$

با کم کردن معادله (21b) از معادله (21a)، نتیجه می شود: $\rightarrow$

$x_2(k) = 0 \qquad m_2 = 0$

و با جایگذاری $x_2(k) = 0$ در معادله (21a) و یا (21b)، عبارت زیر به دست می آید:

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

تذکر(۱): از آنجاییکه $m=(m_1,m_2)$، یک نقطه تعادل معادلات (20a), (20b) است و $e(x_1(k))=0$، آنگاه $\nabla F(x_1(k))=0$.

ویژگی پایداری محلی، حول نقطه تعادل (m_1,m_2) می تواند با استفاده از آنالیزهای سیگنال کوچک، امتحان شود.

فرض کنید سیگنال های آشفته به صورت زیر تعریف شوند:

$$\Psi_1 = x_1 - m_1, \Psi_2 = x_2 - m_2$$

آنگاه معادلات حالت به فرم زیر تبدیل می شوند:

$$(23a)\ \Psi_1(k+1) = \Psi_1(k) - \alpha\nabla F(m_1 + \Psi_1(k)) + \delta\Psi_2(k) + Ce(m_1 + \Psi_1(k))$$

$$(23b)\ \Psi_2(k+1) = -\alpha\nabla F(m_1 + \Psi_1(k)) + \delta\Psi_2(k) + Ce(m_1 + \Psi_1(k))$$

با خطی سازی حول محور تعادل m، معادلات (23a) , (23) به صورت زیر خواهند بود:

$$(24a)\ \Psi_1(k+1) \approx \Psi_1(k) - \alpha\nabla^2 F(m_1)(\Psi_1(k)) + \delta\Psi_2(k) + C\nabla e(m_1)(\Psi_1(k))$$

$$(24b)\ \Psi_2(k+1) \approx -\alpha\nabla^2 F(m_1)(\Psi_1(k)) + \delta\Psi_2(k) + C\nabla e(m_1)(\Psi_1(k))$$

$$\nabla^2 F(m_1) = A \in IR^{Q\times Q}$$

$$\nabla e(m_1) = D \in IR^{Q\times Q}$$

بطوریکه Q، سایز بردار وزن است.

آنگاه :

$$\begin{bmatrix}\Psi_1(k+1)\\ \Psi_2(k+1)\end{bmatrix} + \begin{bmatrix} I-\alpha A + CD & CI\\ -\alpha A + CD & CI\end{bmatrix}\begin{bmatrix}\Psi_1(k)\\ \Psi_2(k)\end{bmatrix} \qquad (25)$$

معادله ماتریسی بالا، می تواند به صورت زیر نوشته شود:

$$(26)\ \Psi(k+1) = \theta\Psi(k)$$

www.kandoocn.com

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

(Ψi) باشد و هر مقدار ویژه (Ψi) در شرط رو به رو صدق کند:

$$|\Psi i| < 1$$

لم (۲):

اگر λ یک مقدار ویژه تابع E باشد، جایی که $E = (\frac{A}{C} - \frac{D}{\alpha})$. آنگاه مقادیر ویژه متناظر با ماتریس θ، توسط جوابهای معادله درجه دوم زیر بدست می آیند:

$$(27)\ \Psi^2 - \Psi(1 + \delta - \alpha\lambda C) + \delta = 0$$

اثبات:

ماتریس θ برای هر $\delta \neq 0,\ D, A$، معکوس پذیر است.فرض کنید Ψ، یک مقدار ویژه θ باشد و آن غیر صفر است (زیرا θ ناویژه[1] است).

فرض کنید که $\underline{Z} = (x,y)$، بردار ویژه غیر صفر متناظر با مقدار ویژه Ψ است. آنگاه:

$$\theta \underline{z} = \Psi \underline{z} \qquad (28)$$

که منجر می شود به:

$$x - \alpha Ax + CDx + \delta y = \Psi x \qquad (29)$$

و

$$-\alpha Ax + CDx + \delta y = \Psi y \qquad (30)$$

با کم کردن معادله (30) از معادله (29)، خواهیم داشت:

$$(31)\ x = \Psi x - \Psi y \rightarrow (\frac{\Psi - 1}{\Psi})x = y$$

با جایگذاری معادله (31) در معادله (30)، خواهیم داشت:

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

با انتخاب $E=\frac{A}{C}-\frac{D}{\alpha}$ و جایگذاری در معادله (32)، رابطه زیر بدست می آید:

$$(33)\; Ex=\left(\frac{(\Psi-1)-\frac{\delta(\Psi-1)}{\Psi}}{-\alpha C}\right)X$$

---

1. Non - singular

از آنجایی که $\left(\frac{(\Psi-1)-\frac{\delta(\Psi-1)}{\Psi}}{-\alpha C}\right)X$، اسکالر و غیر صفر است، اگر بردار x در این معادله صدق کند، x یک بردار ویژه ماتریس E است. آنگاه:

$$Ex = \lambda x \quad (34)$$

با مقایسه معادلات (33), (34)، معادله زیر بدست خواهد آمد:

$$(35)\; \lambda=\left(\frac{(\Psi-1)-\frac{\delta(\Psi-1)}{\Psi}}{-\alpha C}\right)X$$

با مرتب کردن معادله (35)، خواهیم داشت:

$$\Psi^2-\Psi(1+\delta-\alpha\lambda c)+\delta=0 \quad (36)$$

الف. تست پایداری جوری:

تست جوری، برای آنالیز معادلات از هرمرتبه، به کار می رود. معادله (36)، یک معادله درجه دوم است و می توان آن را به صورت $f(z)=a_2 z^2+a_1 z+a_0$ نوشت. شرایط لازم و کافی برای هرچند جمله ای ، بطوری که ریشه ای خارج و روی دایراه واحد نداشته باشد، آن است:

$$|a_0|<a_2 \quad (37)$$

$$f(1)>0 \quad (38)$$

$$(-1)^2 f(-1)>0 \quad (39)$$

www.kandoocn.com www.kandoocn.com www.kandoocn.com www.kandoocn.com

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

$|\delta| < 1$ (40)

$(1+\delta) > ((1+\delta) - \alpha\lambda c)$ (41)

$(1+\delta) > (-(1+\delta) + \alpha\lambda c)$ (42)

از آنجایی که، ضریب ممنتم مثبت است:

$0 < \delta < 1$ (43)

نامعادله (41)، منجر خواهد شد به:

$\alpha\lambda c > 0$ (44)

و نا معادله (42)، به فرم زیر تبدیل خواهد شد:

$$(45)\ \delta > \frac{\alpha\lambda c}{2} - 1$$

با استفاده از نامعادلات (43) و (45)، نتیجه می شود:

$0 < \alpha\lambda c < 4$ (46)

مقادیر α و c، بایستی مثبت باشند تا سیستم یاد بگیرد. بنابراین آن شرط کافی است که همه مقادیر ویژه E، مثبت باشند تا نتیجه بگیریم که m، یک نقطه مینیمم محلی است.

بنابراین همه نقاط مینیمم محلی از F(0)، به طور مجانبی پایدار محلی[۱] هستند.

همچنین از آنجایی که F(0)، یک نقطه مینیمم محلی دارد، ماتریس هسیان[۲] A، ماتریسی مثبت معین[۳] است.

www.kandoocn.com

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

صدق کنند، پایداری سیستم تضمین می شود و سیستم به یک نقطه مینیمم محلی، همگرا خواهد شد.

اگر مقادیر ویژه ماتریس E، نسبتاً بزرگ باشندف ممکن است شرط (46)، نقض شود ولی در اکثر موارد E محدود است، بنابراین اگر α و c، به اندازه کافی کوچک باشند، تمامی نقاط، مینیمم محلی پایدار هستند.

ب. شرط پایداری برای ماتریس D

نامعادله (46)، مقادیر نرخ یادگیری α و ضریب تناسبی c و نیز مقادیر ویژه E را محدود می کنند.

هدف از این بخش آن است که بر روی شرایط کافی پایداری بحث کنیمس به طوری که، رابطه بین پارامترهای α و c رابدست آوریم.

بنابراین باید به بحث روی ماتریس D بپردازیم، به طوری که ماتریس E، مثبت معین باشد:

$$(47)\ E = \frac{A}{C} - \frac{D}{\alpha}$$

تئوری (۱):

E، مثبت معین است، اگر و فقط اگر، ماتریس بالا مثلثی، حقیقی و ناویژه H، به گونه ای یافت شود که $E = H^T H$.

دو حالت را در نظر می گیریم:

حالت اول:

www.kandoocn.com

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

خواهد بود.

و مقادیر ویژه S^TBS، مثبت می باشند. بنابراین مقادیر ویژه معادله (50)، کمتر از $\frac{1}{C}$ نیستند، یعنی:

$$(51)\ \det(S^T ES) \geq \frac{1}{C}$$

---

1. Locally Asymptotically Stable
2. Hessian
3. Positive Definite Matrix
4. Negative Semidefinite Matrix
5. Positive Semidefinite Matrix

بنابراین S^TBS، مثبت معین است، از اینرو:

$$\det (A) = \det (S^{-T}S^{-1}) = (DET (S))^{-2} \qquad (52)$$

بنابراین:

$$(53)\ \frac{1}{C} \leq \det(S^T ES) = (\det(s))^2 \det(E) = \frac{\det(E)}{\det(A)}$$

معادله فوق را می توان به صورت زیر، مرتب کرد:

$$(54)\ C \geq \frac{\det(A)}{\det(E)}$$

حالت دوم:

فرض می کنیم که D، ماتریسی مثبت نیمه معین است.

تئوری (۳):

www.kandoocn.com

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

مثبت معین است و $\frac{I\alpha}{C} > S^T DS$.

اثبات:

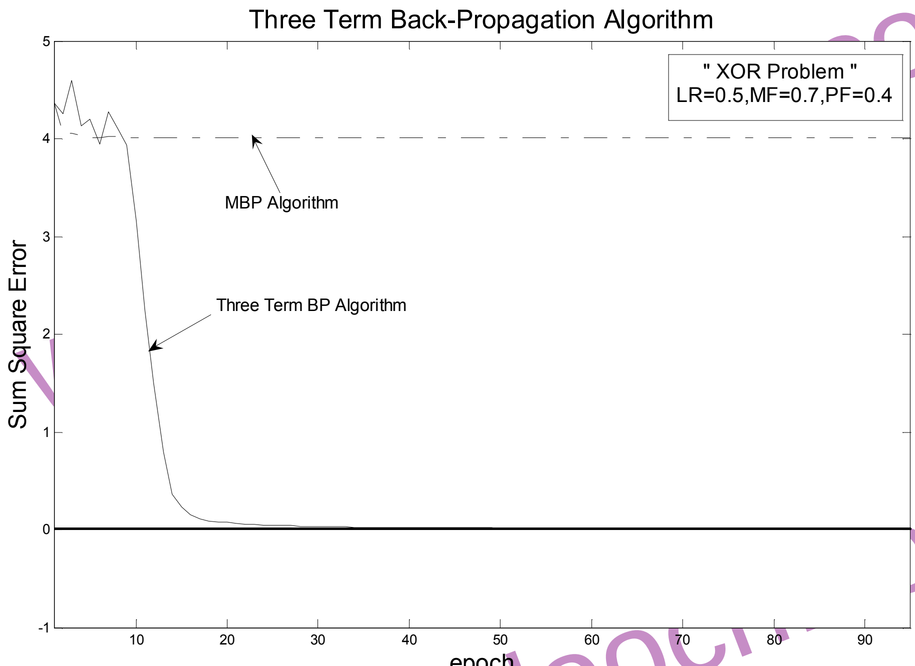
بر اساس اثبات قبلی،

$$(55)\quad S^T ES = \frac{1}{C} - \frac{S^T DS}{\alpha}$$

برای آنکه E، مثبت معین باشد، بایستی طرف راست معادله (55)، مثبت باشد، از اینرو:

$$(56)\quad \frac{I\alpha}{C} > S^T DS$$

از الگوریتم مطرح شده برای جداسازی الگوها در مسأله XOR استفاده شده است. شبیه سازی، نشان می دهد که سرعت همگرایی این الگوریتم، از الگوریتم استانداردپس انتشار خطا بیشتر است.



جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

شکل (۹). منحنی یادگیری الگوریتم BP دارای سه ترم، در مسأله XOR

۳. از رابطه (5)، به خوبی مشخص است که بردارهای حساسیت رفتار شبکه ($\delta_j^L(K)$)، شامل ترم مشتق تابع تحریک نرون می باشد با فرض آنکه تابع تحریک نورونها، در لایه های میانی از نوع تابع زیگموئیدی لگاریتمی می باشند این ترم، $a_j^L(K)(1- a_j^L(K)$ است.

بردارهای حساسیت از لایه آخر به لایه های میانی و سپس به لایه ورودی، برگشت داده می شوند تا پارامترهای شبکه را تصحیح کنند.

زمانی که خروجی $a_j^L(K)$ به مقادیر ۱ و یا 0، میل می کند. مشتق تابع تحریک که دارای فاکتور $a_j^L(K)(1- a_j^L(K)$ می باشد، سبب می شود که بردار حساسیت $\delta_j^L(K)$، بسیار کوچک شود، بنابراین تغییرات ناچیز در پارامتر متناظر با نرون j ام خواهیم داشت.

لذا، فرایند یادگیری و اصلاح پارامترها در الگوریتم BP، بسیار آرام خواهد بود و یا حتی متوقف خواهد شد، حتی اگر مقادیر پارامترهای شبکه، از مقادیر بهینه اشان، فاصله ای زیاد داشته باشند.

بنابراین اصلاح مشتق تابع تحریک f، ضروری است تا بدین ترتیب عملکرد الگوریتم BP را بهبود بخشیم.

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

حساسیت بر روی مشتقات جزئی یا شیب تابع تحریک، بهبودهایی صورت گرفته است.

بردار حساسیت در معادله (5)، به صورت زیر تغییر کرده است:

$$(57)\quad \delta_j^{\,l}(k)=\frac{\delta F(K)}{\delta n_j^{\,l}(k)}=\left[\sum_{m=1}^{s^{l+1}}\delta_m^{l+1}(k)W_{mj}^{l+1}(k)\right]f^{l}(n_j^{l}(k))$$

$$(58)\quad f^{l}(n_j^{l}(k))=\left[a_j^{l}(k)(1-a_j^{l}(k))\right]^{\frac{1}{s}}$$

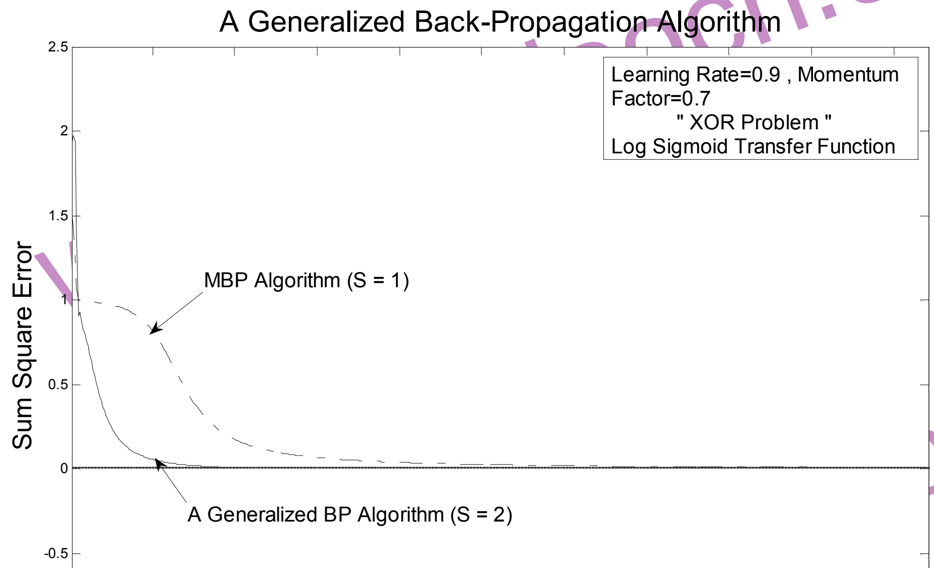
بطوریکه، s یک عدد حقیقی بزرگتر از یک است. $(S\geq 1)$

زمانیکه $S=1$ است، الگوریتم همان الگوریتم استاندارد BP می باشد.

برای $S>1$، بردار حساسیت، زمانی که خروجی به مقدار اشاتباه میل می کند، زیاد خواهد شد. بنابراین نرخ همگرایی فرایند یادگیری می تواند افزایش یابد.

با این وجود، مقدار S، نباید از حد خاصی بیشتر شود زیرا در غیر اینصورت ترم $\left[a_j^{l}(k)(1-a_j^{l}(k))\right]^{\frac{1}{s}}$ به یک میل می کند و بردار حساسیت $\delta_j^{L}(K)$، ناپایدار خواهد شد.

شکل زیر، شبیه سازی که بر اساس این الگوریتم روی مسءله XOR انجام شده است، نشان می دهد. به خوبی مشخص است که سرعت همگرایی این الگوریتم به مراتب بیشتر زا الگوریتم BP می باشد.



جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

شکل (۱۰). منحنی یادگیری الگورتم GBP در مسأله XOR به ازای S=2 و S=1

www.kandoocn.com

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

این الگوریتم، جهت اصلاح مشکل فوق ارائه شده است.

بر اساس این الگوریتم، تنها از علامت مشتق تابع تحریک، جهت اصلاح پارامترهای شبکه استفاده می شود. اندازه مشتق تابع تحریک، هیچ اثری بر تنظیم پارامترهای شبکه ندارد [5], [6].

میزان تغییرات در پارامترهای شبکه، توسط فاکتور delt-inc، افزوده می شود، زمانی که علامت مشتق شاخص اجرایی، نسبت به پارامترهای شبکه دردوتکرار متوالی، تغییر نکند. و زمانی که مشتق خاص اجرایی دردوتکرار متوالی هم علامت نباشند، تغییرات در پارامترهای شبکه توسط فاکتور delt-dec، کاهش می یابد.

الگوریتم Rprop، در زیر خلاصه می گردد.

Rprop Algorithm 1. Choose some small initial value for every update step size Δji(0). 2. Adapt the step size: $if\ \frac{\delta F(k)}{\delta W_{ji}(k)}.\frac{\delta F(k-1)}{\delta W_{ji}(k-1)} \geq 0 \quad \Delta ji(k) = \Delta ji(k-1).(delt.inc)$ $if\ \frac{\delta F(k)}{\delta W_{ji}(k)}.\frac{\delta F(k-1)}{\delta W_{ji}(k-1)} < 0 \quad \Delta ji(k) = \Delta ji(k-1).(delt.dec)$ $\Delta ji(k) = \Delta_{max}, \qquad \Delta ji(k) \geq \Delta_{max}$ $\Delta ji(k) = \Delta_{min}, \qquad \Delta ji(k) \leq \Delta_{minx}$

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

$$if \frac{\delta F(k)}{\delta W_{ji}(k)} > 0 \quad \Delta Wji(k) = -\Delta ji(k)$$

$$if \frac{\delta F(k)}{\delta W_{ji}(k)} < 0 \quad \Delta Wji(k) = \Delta ji(k)$$

$$\Delta Wji(k) = 0 \qquad else$$

مقادیر پیشنهادی برای پارامترها عبارتند از:

$\Delta_{max} = 50$, $\Delta_{min} = 0.000001$, delt- inc = 1.2, delt-dec = 0.5 .

---

1. Resilient Back-Propagation Algorithm

نتیجه گیری

از قانون یادگیری پس انتشار خطا (BP) برای آموزش شبکه ها عصبی چند لایه پیش خور استفاده می شود. با وجود کاربردهای فراوان این الگوریتم یادگیری، هنوز مشکلاتی نیز وجود دارد:

سرعت همگرایی الگوریتم BP، پائین است و ممکن است شبکه به آسانی به نقاط مینیمم محلی همگرا شود. از طرفی، انتخاب نرخ یادگیری، تأثر بسزایی در سرعت همگرایی آموزش شبکه عصبی دارند.

در این گزارش، الگوریتم های جدیدی، جهت بهبود الگوریتم BP، ارائه شده است.

برخی از این روش ها بر مبنای نرخ یادگیری تطبیقی می باشند. بدین صورت که نرخ یادگیری به هنگام پروسه آموزش تغییر می کند تا عملکرد در الگوریتم BP استاندارد بهبود بخشیده شود، نرخ یادگیری تطبیقی سعی می کند که نرخ یادگیری را تا آنجایی که ممکن است و سیستم ناپایدار نشده است، افزایش دهد.

www.kandoocn.com

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

ترم است. در این الگوریتم، ترم جدیدی به نام ضریب تناسبی (PE)، علاوه بر دوترم نرخ یادگیری و ضریب ممنتم، به الگوریتم BP، اضافه شده است.

الگوریتم مطرح شده، همانند کنترل کننده PID، عمل می کند.

همچنین در این گزارش، به بررسی و آنالیز پایداری الگوریتم مطرح شده، پرداخته شده است.

آنالیز پایداری به دلیل آن است که، شرایطی را که باید پارامترهای یادگیری در آن صدق کنند، تا الگوریتم پایدار بماند، بدست آوریم.

در آخر نیز، الگوریتم هایی ارائه شده است، یکی از این الگوریتم ها، الگوریتم پس انتشار خطای بهبود پذیر (Rprop) است. در این الگوریتم، تنها از علامت مشتق تابع تحریک نسب به پارامترهای شبکه، جهت تنظیم پارامترهای شبکه استفاده می شود و اندازه مشتق تابع تحریک، هیچ تأثیری بر تنظیم پارامترهای شبکه ندارد.

در الگوریتم دیگر، جهت افزایش سیگنال حساسیت، اصلاحاتی بر روی مشتق تابع تحریک نرونها، انجام گرفته است.

شبیه سازی انجام شده نشان می دهد، که سرعت همگرایی الگوریتم های مطرح شده نسبت به الگوریتم استاندارد BP، بیشتر است. از طرفی الگوریتم های مطرح شده، محاسبات پیچیده ای ندارد و از همان سادگی الگوریتم BP، برخوردار می باشند.

www.kandoocn.com

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

[1] Wen Jin-Wei, Zhao Jia-Li, Luo Si-Wei and Han Zhen, "The Improvements of BP Neural Network Learninig Algorithm", IEEE, 2000.

[2] Chien-Cheng Yu, and Bin-Da Liu, "A BACKPROPAGATION ALGORITHM WITH ADAPTIVE LEARNING RATE AND MOMENTOM COEFFI CIENT", IEEE, 2000

[3] Yahya H.Zweiri, James F.Whidborne, Kaspar Althoefer and Lakmal D.Seneviratne, "A New Three-Term Backpropagation Algorithm with Convergence Analysis", Proceeding of the IEEE International Conference on Robotics 8Automation, 2002.

[4] S.C.Ng, S.H.Leung and A.Luk, "A Generalized Back-Propagtion Algorithm for Faster Convergence", IEEE, 1996.

[5] Riedmiller.M and H.Braun, "A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm", proceeding of the IEEE International Conference on Neural Networks, 1993.

[6] Z.Zainuddin, N. Mahat and Y.Abu Hassan, "Improving the Convergence of the Backpropagation Algorithm USING Local Adaptive Techniques", Tnternational Journal of Computational Intelligence, Vol. 1, No. 3, 2004.

جهت خرید فایل word به سایت www.kandoocn.com مراجعه کنید و یا باشماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

Learning Algorithm", IEEE, 1992.

[8] Neural Network Toolbox.

مرجع فارسی:

۱- مبانی شبکه های عصبی (هوش محاسباتی)، دکتر محمد باقر منهاج، انتشارات دانشگاه صنعتی امیر کبیر، چاپ دومف پائیز ۱۳۸۱.

www.kandoocn.com