

فیلتر غیرخطی تطبیقی عصبی شناساگر با نرخ آموزش بهینه برای همگرایی پارامترها بر پایه گرادیان نزولی

فاطمه علی بخشی^۱، محمد تشنه لب^۲، مهدی علی بخشی^۳، محمد منثوری^۴

۱- کارشناسی ارشد، دانشکده فنی و مهندسی - دانشگاه آزاد اسلامی واحد تهران جنوب - تهران - ایران

alibakhshi.fatemeh@gmail.com

۲- استاد گروه کنترل، دانشکده برق و کامپیوتر - دانشگاه صنعتی خواجه نصیرالدین طوسی - تهران - ایران

teshnehlab@eetd.kntu.ac.ir

۳- کارشناسی ارشد، دانشگاه آزاد اسلامی واحد بروجرد - باشگاه پژوهشگران جوان و نخبگان - بروجرد -

ایران

mehdi.alibakhshi@yahoo.com

۴- دانشجوی دکتری، دانشکده برق و کامپیوتر - دانشگاه صنعتی خواجه نصیرالدین طوسی - تهران - ایران

mohammad.mansouri@ee.kntu.ac.ir

چکیده: بحث همگرایی در شبکه‌های عصبی شناساگر و کنترل‌کننده یکی از موارد پراهمیت در مهندسی کنترل می‌باشد، که در این راستا تحقیقات متنوعی صورت پذیرفته است که عمدتاً در چهار چوب شبکه‌های عصبی معمولی بوده است. در این مقاله الگوریتم گرادیان نزولی تطبیقی با یادگیری پارامترهای شبکه عصبی چند لایه بصورت بهنگام بر اساس بسط سری تیلور خطای خروجی مطرح و همگرایی الگوریتم آموزشی بررسی شده است. برای افزایش دقت و سرعت همگرایی، جمله‌های مرتبه دوم و بالاتر بسط سری تیلور با استفاده از گرادیان نزولی بروز رسانی می‌شود. الگوریتم آموزشی معرفی شده برای دو مثال شبیه‌سازی شده است. نتایج شبیه‌سازی کاهش خطا و افزایش سرعت همگرایی را نشان می‌دهند.

واژه‌های کلیدی: الگوریتم گرادیان نزولی، بسط سری تیلور، شبکه عصبی، همگرایی.

۱- مقدمه

به شمار می‌رود [۴]. پرکاربردترین معماری شبکه عصبی مصنوعی، چند لایه پیشرو است. مشهورترین روش آموزش شبکه عصبی پیشرو چند لایه الگوریتم پس‌انتشار^۱ بر اساس روش گرادیان نزولی^۲ است. برای این الگوریتم تعیین نرخ یادگیری مناسب که فرآیند آموزش را پایدار سازد ضروری است. نرخ یادگیری بزرگ ممکن است سیستم را ناپایدار کند. برای اطمینان از یادگیری پایدار باید نرخ یادگیری به اندازه کافی کوچک باشد. با نرخ آموزش کوچک نیز ممکن است زمان آموزش طولانی شود [۲۰]. مطالب بیان شده، مشکلات ذاتی الگوریتم پس‌انتشار بر اساس گرادیان نزولی است [۲۸،۲۰]. ویژگی‌های همگرایی الگوریتم پس‌انتشار بر

در دو دهه اخیر غالباً در تجارت و صنعت سخت افزار و نرم افزار از روش‌های محاسبات نرم استفاده می‌شود. شبکه عصبی مصنوعی یکی از اجزاء مهم هر محاسبات نرم

^۱ تاریخ ارسال مقاله : ۱۳۹۲/۹/۲۸

تاریخ پذیرش مقاله : ۱۳۹۳/۱۱/۲۶

نام نویسنده مسئول : فاطمه علی بخشی

نشانی نویسنده مسئول : ایران - تهران - دانشگاه آزاد اسلامی

واحد تهران جنوب - دانشکده فنی و مهندسی

فیلتر غیرخطی تطبیقی عصبی شناساگر با نرخ آموزش بهینه برای همگرایی پارامترها بر پایه گرادیان نزولی

برای بهبود الگوریتم گرادیان نزولی غیرخطی نرمال، الگوریتم گرادیان نزولی غیرخطی نرمال تطبیقی کامل^۷ ارائه شده و نشان داده شده است که سرعت همگرایی بهبود یافته است [۱۵].

در این مقاله برای شناسایی سیستم‌های غیرخطی، شبکه عصبی چند لایه با آموزش بهنگام استفاده شده است. نرخ یادگیری شبکه عصبی چندلایه براساس بسط سری تیلور خطای خروجی حاصل شده است. برای جمله‌های مرتبه دوم و بالاتر بسط سری تیلور عدد ثابتی با توجه به شرایط سیستم فرض شده است. در ادامه برای افزایش سرعت همگرایی جمله‌های مرتبه دوم و بالاتر بسط سری تیلور نیز بصورت تطبیقی با استفاده از گرادیان نزولی بروز رسانی شده است.

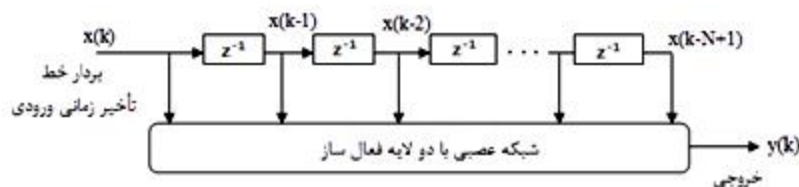
در بخش دوم این مقاله به الگوریتم آموزشی بکار برده شده در شبکه عصبی MLP پرداخته می‌شود و در بخش سوم همگرایی نرخ‌های آموزش مورد بحث قرار گرفته است. در بخش چهارم نتایج شبیه‌سازی انجام شده برای دو سیستم غیرخطی ملاحظه می‌شود و در بخش پنجم به بحث و نتیجه‌گیری پرداخته شده است.

۲- الگوریتم آموزش

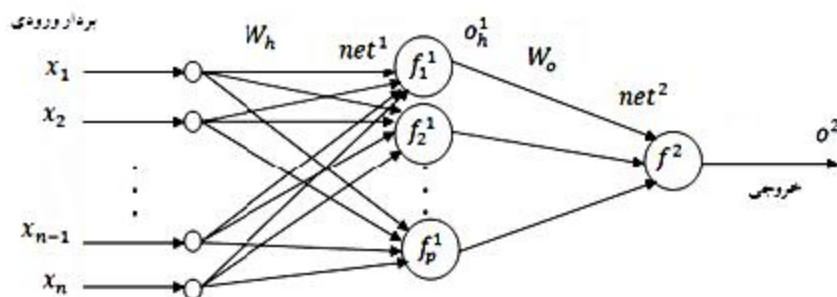
سیگنال‌های بلادرنگ^۸ اغلب متغیر و دارای مدل‌های ریاضی غیرخطی و مشتقات پیچیده هستند [۱۶، ۱۵]. با کمک فیلترهای تطبیقی غیرخطی و شبکه عصبی مدل‌های ساده‌تر را می‌توان جایگزین این مدل‌های پیچیده کرد [۱۷، ۱۵۸]. فیلترهای تطبیقی همگرایی آهسته دارند [۱۸، ۱۵، ۱۰] و برای بهبود سرعت همگرایی از شبکه عصبی استفاده می‌شود. در این مقاله از شبکه عصبی MLP با دو لایه فعال‌ساز بصورت نشان داده شده در شکل (۱) استفاده می‌شود. جزئیات ساختار فیلتر تطبیقی عصبی غیرخطی با دو لایه فعال‌ساز شکل (۱) در شکل (۲) نشان داده شده است.

اساس گرادیان نزولی در [۲۹، ۲۴، ۲۲، ۱۲، ۷، ۵] و [۲۰] بحث شده است. قوانین یادگیری می‌تواند بصورت یک روش برای اصلاح و یا تغییر پارامترهای شبکه عصبی برای آموزش شبکه به منظور انجام برخی کارها تعریف شود [۱۱، ۶]. به هر حال از مشکل ذاتی در این الگوریتم‌ها، همگرایی به مینیمم محلی و نرخ سرعت هستند که به روند یادگیری حساس می‌باشند [۳۰، ۲۱]. شبکه‌های عصبی پیشرو با الگوریتم آموزشی پس‌انتشار بطورگسترده‌ای در بازشناسایی الگو، بهینه‌سازی، مدل‌سازی، دسته‌بندی، شناسایی و کنترل کاربرد دارند [۱۳] و [۳۱، ۳۲]. رابطه سیستم‌های غیرخطی پیچیده با آموزش شبکه عصبی می‌تواند تقریب زده شود [۲۳]. با این حال نرخ همگرایی کوچک، مسئله مینیمم محلی و عدم پایداری از جمله مسائلی است که سیستم‌های بر پایه این الگوریتم با آن رو به رو هستند.

در دهه‌های اخیر تلاش‌های بسیاری برای بهبود همگرایی پس‌انتشار انجام شده است. در [۲۷، ۲۶، ۲۵، ۹] برای بهبود الگوریتم پس‌انتشار از آموزش بهنگام استفاده شده است. یادگیری بر پایه الگوریتم‌های GD شامل یادگیری بازگشتی زمان واقعی^۳، مشتق‌های مرتبه‌های بالاتر و غیره می‌باشد [۱]. روش‌های بر پایه مشتق سرعت همگرایی را بهبود بخشیده است، اما این روش‌ها تمایل به همگرایی به مینیمم محلی دارند [۲، ۱۹]. علاوه بر این با توجه به وابستگی آن‌ها به تحلیل مشتقات، آن‌ها را به تابع‌های هدف خاص محدود می‌کند [۲]. برای رفع مشکلاتی مانند سرعت همگرایی پایین و محاسبات سنگین در بروز رسانی پارامترها در [۲۱] الگوریتم آموزشی گرادیان نزولی وفقی مقاوم^۴ برای سیستم‌های چند ورودی و تک خروجی طراحی شده است. با توسعه الگوریتم گرادیان نزولی تطبیقی مقاوم، الگوریتم گرادیان نزولی وفقی مقاوم چند متغیره^۵ در [۳۰] ارائه شده است. الگوریتم گرادیان نزولی غیرخطی نرمال^۶ برای بهبود همگرایی در فیلترهای غیرخطی FIR در [۱۴] مورد بررسی قرار گرفته است.



شکل (۱): فیلتر تطبیقی عصبی غیر خطی



شکل (۲) : جزئیات ساختار شبکه عصبی با دو لایه فعال ساز

در روابط فوق

وزن‌های لایه فعال‌ساز خروجی شبکه عصبی می‌باشد که در آن p تعداد نرون‌های لایه فعال‌ساز میانی است و

ماتریس وزن‌های لایه فعال‌ساز میانی شبکه عصبی است. $\eta^o(k)$ نرخ آموزش پارامترهای لایه فعال‌ساز خروجی و $\eta^h(k)$ نرخ آموزش پارامترهای لایه فعال‌ساز میانی است و خطا بصورت $E(k) = \frac{1}{2}e^2(k)$ می‌باشد. در ادامه به تشریح الگوریتم آموزش پرداخته می‌شود.

۲-۱- ارائه الگوریتم گرادیان نزولی غیرخطی

نرمال برای شبکه MLP

اگر تابع هزینه بصورت زیر تعریف شود:

$$E(k) = \frac{1}{2}e^2(k) \quad (۴)$$

با استفاده از بسط سری تیلور خطای خروجی و رابطه

(۱) خواهیم داشت:

ساختار شبکه عصبی دارای سه لایه می‌باشد. با فرض اینکه پارامترهای شبکه عصبی در زمان k تغییر می‌کند، سیگنال‌های ورودی می‌باشد که در آن n تعداد ورودی‌هاست، لذا مطابق شکل (۲) خواهیم داشت:

$$e(k) = d(k) - y(k) = d(k) - O^2(k) = d(k) - f^2(O_h^1(k)W_o(k)) \quad (۱)$$

در رابطه فوق $e(k)$ خطای خروجی واقعی، $d(k)$ خروجی مطلوب شبکه عصبی و $O^2(k)$ خروجی نهایی شبکه عصبی و $O_h^1(k)$ خروجی تک تک نرون‌های لایه فعال‌ساز میانی شبکه عصبی است. $f^1(\cdot)$ و $f^2(\cdot)$ توابع تحریک غیرخطی لایه‌های فعال‌ساز هستند. بر اساس شکل (۲)، $net^2(k) = O_h^1(k)W_o(k)$ و $net^1(k) = X(k)W_h(k)$ می‌باشند. حال با استفاده از روش گرادیان نزولی قانون تنظیم وزن‌ها در لایه‌های فعال‌ساز خروجی و میانی به ترتیب زیر محاسبه می‌شوند:

$$W_o(k+1) = W_o(k) + \eta^o(k) \left(-\frac{\partial E(k)}{\partial W_o(k)} \right) \quad (۲)$$

$$W_h(k+1) = W_h(k) + \eta^h(k) \left(-\frac{\partial E(k)}{\partial W_h(k)} \right) \quad (۳)$$

فیلتر غیرخطی تطبیقی عصبی شناساگر با نرخ آموزش بهینه برای همگرایی پارامترها بر پایه گرادیان نزولی

که متغیر c_o مقداری ثابت است و نشان دهنده جمله‌های مرتبه دوم و بالاتر بسط سری تیلور رابطه (۵) است. از گرادیان تابع هزینه نسبت $W_h(k)$ خواهیم داشت:

$$\frac{\partial E(k)}{\partial W_h(k)} = \left(\frac{\partial E(k)}{\partial e(k)} \right) \times \left(\frac{\partial e(k)}{\partial O^2(k)} \right) \times \left(\frac{\partial O^2(k)}{\partial net^2(k)} \right) \times \left(\frac{\partial net^2(k)}{\partial O_h^1(k)} \right) \times \left(\frac{\partial O_h^1(k)}{\partial net^1(k)} \right) \times \left(\frac{\partial net^1(k)}{\partial W_h(k)} \right) \quad (12)$$

$= -e(k) f^{2'}(k) W_o(k) f^{1'}(k) X(k)$
با جایگذاری این مقدار در رابطه (۳) نتیجه زیر حاصل می‌شود:

$$W_h(k+1) = W_h(k) + \eta^h(k) e(k) f^{2'}(k) f^{1'}(k) W_o(k) X(k) \quad (13)$$

با استفاده از بسط سری تیلور خطای خروجی و رابطه (۱) خواهیم داشت:

$$e(k+1) = e(k) + \sum_{i=1}^n \frac{\partial e(k)}{\partial w_{hi}(k)} \Delta w_{hi}(k) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 e(k)}{\partial w_{hi}(k) \partial w_{hj}(k)} \Delta w_{hi}(k) \Delta w_{hj}(k) + h.o.t \quad (14)$$

که $h.o.t$ نشان دهنده جمله‌های مرتبه بالاتر بسط سری تیلور می‌باشد. از روابط (۱) و (۱۳) خواهیم داشت:

$$\frac{\partial e(k)}{\partial W_{hi}(k)} = -f^{2'}(k) f^{1'}(k) W_o(k) X(k) \quad i = 1, 2, \dots, n \quad (15)$$

$$\Delta w_{hi}(k) = \eta^h(k) e(k) f^{2'}(k) f^{1'}(k) W_o(k) X(k) \quad i = 1, 2, \dots, n \quad (16)$$

با جایگذاری روابط (۱۵) و (۱۶) در رابطه (۱۴) خواهیم داشت:

$$e(k+1) = e(k) \left[1 - \eta^h(k) (f^{2'}(k))^2 (f^{1'}(k))^2 \|W_o(k)\|^2 \|X(k)\|^2 \right] \quad (17)$$

$$e(k+1) = e(k) + \sum_{i=1}^p \frac{\partial e(k)}{\partial w_{oi}(k)} \Delta w_{oi}(k) + \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p \frac{\partial^2 e(k)}{\partial w_{oi}(k) \partial w_{oj}(k)} \Delta w_{oi}(k) \Delta w_{oj}(k) + h.o.t \quad (5)$$

که $h.o.t$ نشان دهنده جمله‌های مرتبه بالاتر بسط سری تیلور است. با استفاده از گرادیان نزولی خطا نسبت $W_o(k)$ خواهیم داشت:

$$\frac{\partial E(k)}{\partial W_o(k)} = \left(\frac{\partial E(k)}{\partial e(k)} \right) \times \left(\frac{\partial e(k)}{\partial O^2(k)} \right) \times \left(\frac{\partial O^2(k)}{\partial net^2(k)} \right) \times \left(\frac{\partial net^2(k)}{\partial W_o(k)} \right) = -e(k) f^{2'}(k) O_h^1(k) \quad (6)$$

با توجه به رابطه (۲) و (۶) داریم:

$$W_o(k+1) = W_o(k) + \eta^o(k) e(k) f^{2'}(k) O_h^1(k) \quad (7)$$

از روابط (۱) و (۷) خواهیم داشت:

$$\frac{\partial e(k)}{\partial w_{oi}(k)} = -f^{2'}(k) O_h^1(k) \quad i = 1, 2, \dots, p \quad (8)$$

با جایگزین کردن روابط (۸) و (۹) در رابطه (۵) و ساده‌سازی نتیجه زیر حاصل می‌شود:

$$\Delta w_{oi}(k) = \eta^o(k) e(k) f^{2'}(k) O_h^1(k) \quad i = 1, 2, \dots, p \quad (9)$$

$$e(k+1) = e(k) \left[1 - \eta^o(k) (f^{2'}(k))^2 \|O_h^1(k)\|^2 \right] \quad (10)$$

که در رابطه فوق $\| \cdot \|$ نرم اقلیدسی^۹ است. با حل رابطه (۱۰) و با توجه به اینکه اگر خطا در دو گام متوالی برابر شوند، نرخ آموزش پارامترهای لایه فعال‌ساز خروجی بصورت زیر حاصل می‌شود:

$$\eta^o(k) = \frac{1}{(f^{2'}(k))^2 \|O_h^1(k)\|^2 + c_o} \quad (11)$$

که $E(k) = \nabla_{C_o(k-1)} \left[\frac{1}{2} e^2(k) \right]$ گرادیان تابع هزینه

نسبت به متغیر $C_o(k-1)$ است و ρ اندازه گام این الگوریتم تطبیقی است. همانطور که بیان شد، مقدار $C_o(k)$ باقیمانده تطبیقی از بسط سری تیلور است در حالی که $k > 0$ می باشد. با توجه به اینکه جمله های مرتبه های بالاتر بسط سری تیلور به مراتب کوچکتر از جمله اول بسط سری تیلور می باشد می توان مقدار اولیه $C_o(0)$ را با توجه به نوع سیستم غیرخطی در بازه $[0,1]$ انتخاب نمود.

از رابطه (۲۱)، معادله تطبیقی $C_o(k)$ بصورت رابطه زیر بدست می آید:

$$C_o(k) = C_o(k-1)$$

$$-\rho \frac{f^{2'}(k)f^{2'}(k-1)O_h^1(k)O_h^1(k-1)e(k)e(k-1)}{\left((f^{2'}(k-1))^2 \|O_h^1(k-1)\|^2 + C_o(k-1) \right)^2} \quad (22)$$

از ترکیب روابط (۱۱) و (۲۲) نرخ آموزش لایه فعال ساز خروجی بصورت رابطه زیر بدست می آید:

$$\eta^o(k) = \frac{1}{(f^{2'}(k))^2 \|O_h^1(k)\|^2 + C_o(k)} \quad (23)$$

که در آن $C_o(k)$ از رابطه (۲۲) حاصل می شود. همانند الگوریتمی که برای $C_o(k)$ بدست آورده شد، برای تنظیم $C_h(k)$ خواهیم داشت:

$$C_h(k) = C_h(k-1) - \rho \nabla_{C_h(k-1)} \left[\frac{1}{2} e^2(k) \right] \quad (24)$$

که $E(k) = \nabla_{C_h(k-1)} \left[\frac{1}{2} e^2(k) \right]$ گرادیان تابع هزینه

نسبت به متغیر $C_h(k-1)$ است و ρ اندازه گام این الگوریتم است. مقدار باقیمانده تطبیقی از بسط سری تیلور در حالی که $k > 0$ است. با توجه به اینکه جمله های مرتبه های بالاتر بسط سری تیلور به مراتب کوچکتر از جمله اول بسط سری تیلور می باشد می توان مقدار اولیه $C_o(k)$ را با توجه به نوع سیستم غیرخطی در بازه $[0,1]$ انتخاب نمود.

از رابطه (۲۴)، معادله تطبیقی $C_h(k)$ بصورت رابطه زیر حاصل شده است:

با حل رابطه (۱۷) نرخ آموزش پارامترهای لایه فعال ساز

میانی بصورت زیر بدست می آید:

$$\eta^h(k) = \frac{1}{(f^{2'}(k))^2 (f^{1'}(k))^2 \|W_o(k)\|^2 \|X(k)\|^2 + C_h} \quad (18)$$

که متغیر C_h مقداری ثابت است و نشان دهنده جمله های مرتبه دوم و بالاتر بسط سری تیلور رابطه (۱۴)، و مربوط به تنظیم پارامترهای لایه فعال ساز میانی است.

در روابط (۱۱) و (۱۸) برای جمله های مرتبه دوم و بالاتر بسط سری تیلور خطای خروجی متغیر ناشناخته ای در نظر گرفته شده است و هر چند که C_o و C_h متغیر با زمان هستند، برای سادگی این مقادیر ثابت فرض شده اند. اما برای سیگنال های ورودی غیرخطی و متغیر، فرض کردن C_o و C_h بصورت مقادیر ثابت مناسب نمی باشند، بنابراین این مقادیر را بصورت متغیر $C_o(k)$ و $C_h(k)$ در نظر می گیریم. لذا نیاز داریم که $C_o(k)$ و $C_h(k)$ با استفاده از یک روش گرادیان نزولی و متناسب با تغییر سیگنال های ورودی، در گام های مختلف تنظیم شوند.

۲-۲-۲- ارائه الگوریتم گرادیان نزولی غیرخطی

نرمال تطبیقی کامل برای شبکه MLP

با جایگزین کردن نرخ های آموزش بدست آمده در

رابطه تنظیم وزن های لایه های فعال ساز خروجی و میانی خواهیم داشت:

$$W_o(k+1) = W_o(k) + \frac{e(k)f^{2'}(k)O_h^1(k)}{(f^{2'}(k))^2 \|O_h^1(k)\|^2 + C_o(k)} \quad (19)$$

$$W_h(k+1) = W_h(k) + \frac{e(k)f^{2'}(k)f^{1'}(k)W_o(k)X(k)}{(f^{2'}(k))^2 (f^{1'}(k))^2 \|W_o(k)\|^2 \|X(k)\|^2 + C_h(k)} \quad (20)$$

تنظیم تطبیقی باقیمانده مرتبه دوم و بالاتر بسط سری

تیلور بر اساس گرادیان بصورت زیر تعریف می شود:

$$C_o(k) = C_o(k-1) - \rho \nabla_{C_o(k-1)} \left[\frac{1}{2} e^2(k) \right] \quad (21)$$

فیلتر غیرخطی تطبیقی عصبی شناساگر با نرخ آموزش بهینه برای همگرایی پارامترها بر پایه گرادیان نزولی

$$C_h(k) = C_h(k-1) - \rho \frac{f^{2'}(k)f^{2'}(k-1)f^{1'}(k)f^{1'}(k-1)W_o(k)W_o(k-1)X(k)X(k-1)e(k)e(k-1)}{\left((f^{2'}(k-1))^2 (f^{1'}(k-1))^2 \|W_o(k-1)\|^2 \|X(k-1)\|^2 + C_h(k-1) \right)^2} \quad (25)$$

از روابط (۱۸) و (۲۴) نرخ آموزش لایه فعال‌ساز میانی بصورت رابطه زیر حاصل می‌شود:

$$\eta^h(k) = \frac{1}{(f^{2'}(k))^2 (f^{1'}(k))^2 \|W_o(k)\|^2 \|X(k)\|^2 + C_h(k)} \quad (26)$$

که در آن $C_h(k)$ از رابطه (۲۵) حاصل می‌شود.

$$0 < \frac{1}{(f^{2'}(k))^2 \|O_h^1(k)\|^2 + C_o(k)} < \frac{2}{(f^{2'}(k))^2 \|O_h^1(k)\|^2} \quad (30)$$

با حل نامساوی فوق برای $C_o(k)$ رابطه زیر بدست می‌آید:

$$\frac{-(f^{2'}(k))^2 \|O_h^1(k)\|^2}{2} < C_o(k) \quad (31)$$

رابطه (۳۱) کران پایین $C_o(k)$ می‌باشد، بطوریکه رابطه (۲۷) برقرار باشد. همچنین این الگوریتم برای مقادیر $|e(k)| \rightarrow 0$ و $\eta^h(k)$ باید همگرا شود، یعنی $|e(k)| \rightarrow 0$ وقتی $k \rightarrow \infty$ سوق پیدا می‌کند، لذا با توجه به رابطه (۱۷) خواهیم داشت:

$$|e(k+1)| \leq \left| 1 - \eta^h(k) (f^{2'}(k))^2 (f^{1'}(k))^2 \|W_o(k)\|^2 \|X(k)\|^2 \right| |e(k)| \quad (32)$$

از این رو بدست می‌آید:

$$\left| 1 - \eta^h(k) (f^{2'}(k))^2 (f^{1'}(k))^2 \|W_o(k)\|^2 \|X(k)\|^2 \right| < 1 \quad (33)$$

و با حل نامساوی رابطه (۳۳) خواهیم داشت:

$$0 < \eta^h(k) < \frac{2}{(f^{2'}(k))^2 (f^{1'}(k))^2 \|W_o(k)\|^2 \|X(k)\|^2} \quad (34)$$

رابطه (۳۴) محدوده همگرایی $\eta^h(k)$ را مشخص می‌کند. برای تعیین محدوده همگرایی $C_h(k)$ ، رابطه (۲۶)

۳- بررسی همگرایی

با تعیین نرخ یادگیری توسط الگوریتم آموزش، این الگوریتم برای مقادیر $\eta^o(k)$ ، $C_o(k)$ ، $\eta^h(k)$ و $C_h(k)$ باید همگرا شود. بطوریکه $|e(k)| \rightarrow 0$ وقتی $k \rightarrow \infty$ سوق پیدا می‌کند. بنابراین برای همگرایی یکنواخت باید خطا در یک لحظه بعد کوچکتر و یا اینکه مساوی خطای لحظه حال حاضر شود و یا به عبارتی $|e(k+1)|$ کوچکتر مساوی $|e(k)|$ باشد. با توجه به رابطه (۱۰) که از بسط سری تیلور برای خطا بدست آمده است، می‌توان نوشت:

$$|e(k+1)| \leq \left| 1 - \eta^o(k) (f^{2'}(k))^2 \|O_h^1(k)\|^2 \right| |e(k)| \quad (27)$$

از این رو با فرض اینکه دو طرف نامعادله فوق مساوی باشد:

$$\left| 1 - \eta^o(k) (f^{2'}(k))^2 \|O_h^1(k)\|^2 \right| < 1 \quad (28)$$

با توجه به رابطه ریاضی $|a| < b \Rightarrow -b < a < b$ خواهیم داشت:

$$0 < \eta^o(k) < \frac{2}{(f^{2'}(k))^2 \|O_h^1(k)\|^2} \quad (29)$$

که این رابطه محدوده همگرایی $\eta^o(k)$ را مشخص می‌کند. برای تعیین محدوده همگرایی $C_o(k)$ ، رابطه (۲۳) را در رابطه نامساوی (۲۹) جایگذاری می‌کنیم:

مثال اول: شبکه عصبی را برای شناسایی اجزای غیرخطی در یک سیستم کنترل بهنگام بکار می‌بریم و به ارزیابی الگوریتم پیشنهاد شده می‌پردازیم. معادله سیستم به صورت زیر می‌باشد [۳]:

$$y(k+1) = f(y(k), y(k-1), y(k-2), u(k), u(k-1)) \quad (37)$$

که در آن $u(k)$ و $y(k)$ به ترتیب ورودی و خروجی سیستم در زمان k می‌باشند و تابع ناشناخته $f(\cdot)$ بصورت زیر می‌باشد:

$$f(x_1, x_2, x_3, x_4, x_5) = \frac{x_1 x_2 x_3 x_5 (x_3 - 1) + x_4}{1 + x_2^2 + x_3^3} \quad (38)$$

برای شبیه‌سازی این سیستم غیرخطی از شبکه عصبی با ساختار نشان داده شده در شکل (۲) استفاده شده که در آن n برابر ۵ و p برابر ۱۵ فرض شده است. در این شبکه عصبی فرض شده که f^2 تابع خطی و f^1 تابع سیگموئید دو قطبی است که بصورت رابطه زیر تعریف می‌شود:

$$f^1(net) = \frac{1 - e^{-net}}{1 + e^{-net}} \quad (39)$$

که در آن net مجموع وزندار شده ورودی‌ها است. برای مقایسه از معیار میانگین مجموع مربعات خطا^۱ استفاده شده است. در شکل‌های (۴)، (۵) و (۶) مقدار ثابت برای باقیمانده بسط سری تیلور خطای خروجی در نظر گرفته شده است. شکل (۴) همگرایی پارامترهای الگوریتم آموزشی را نشان می‌دهد.

در شکل (۵)، نرخ‌های آموزش برابر با حد بالای قیود بدست آمده در بخش ۳ انتخاب شده است.

شکل (۶) واگرایی الگوریتم آموزشی در صورت عدم رعایت قیود بدست آمده در بخش ۳ را نشان می‌دهد.

تنظیم تطبیقی باقیمانده بسط سری تیلور بر اساس گرادیان نزولی با رعایت شرایط همگرایی در شکل (۷) نشان داده شده است.

را در رابطه نامساوی (۳۴) جایگذاری می‌کنیم:

$$0 < \frac{1}{(f^{2'}(k))^2 (f^{1'}(k))^2 \|W_o(k)\|^2 \|X(k)\|^2 + C_h(k)} < \frac{2}{(f^{2'}(k))^2 (f^{1'}(k))^2 \|W_o(k)\|^2 \|X(k)\|^2} \quad (35)$$

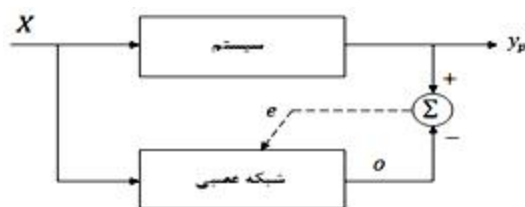
با حل نامساوی رابطه (۳۵) برای $C_h(k)$ ، رابطه زیر بدست می‌آید:

$$\frac{-(f^{2'}(k))^2 (f^{1'}(k))^2 \|W_o(k)\|^2 \|X(k)\|^2}{2} < C_h(k) \quad (36)$$

رابطه (۳۶) کران پایین $C_h(k)$ می‌باشد، بطوریکه رابطه (۳۲) برقرار باشد.

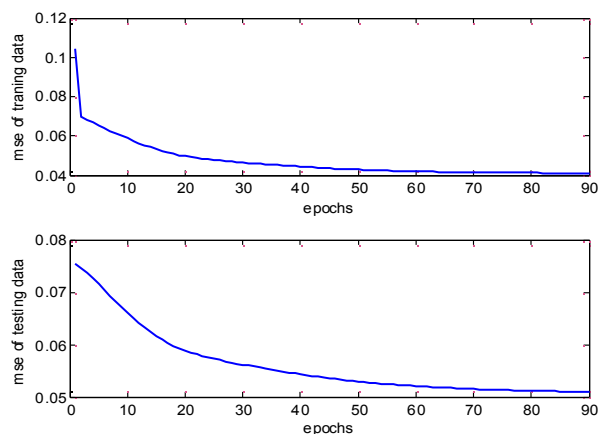
۴- نتایج شبیه‌سازی

در این بخش الگوریتم‌های پیشنهاد شده در بخش‌های ۲-۲ و ۲-۱ بر روی سیستم غیرخطی در قالب دو مثال شبیه‌سازی شده است. در هر دو مثال از ۱۰۰۰ داده تصادفی استفاده شده است. با تعداد داده‌های آموزشی کم، مقدار خطا کمی افزایش می‌یابد. اما با این حال می‌توان از این الگوریتم آموزشی برای شبکه‌های عصبی با تعداد داده‌های کم نیز استفاده نمود. داده‌های تصادفی در مثال‌ها به داده‌های آموزش و داده‌های تست تقسیم شده‌اند و شبکه عصبی به عنوان شناساگر سیستم غیرخطی مورد استفاده قرار گرفته است. این ساختار در شکل (۳) نشان داده شده است که در این شکل y_p خروجی واقعی سیستم و O خروجی شبکه عصبی می‌باشد.

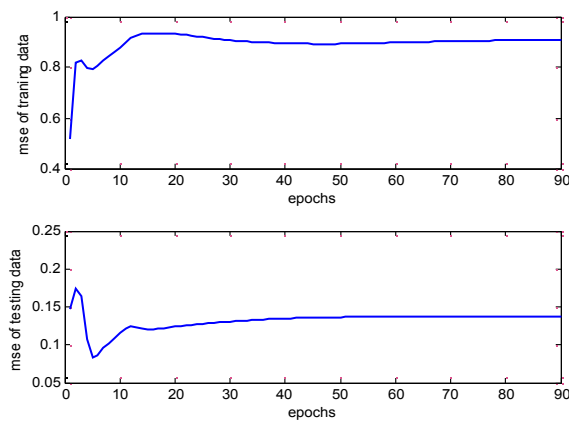


شکل (۳): شبکه عصبی به عنوان شناساگر غیرخطی

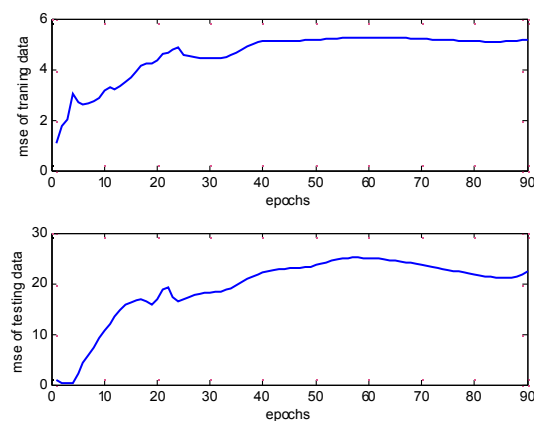
فیلتر غیرخطی تطبیقی عصبی شناساگر با نرخ آموزش بهینه برای همگرایی پارامترها بر پایه گرادیان نزولی



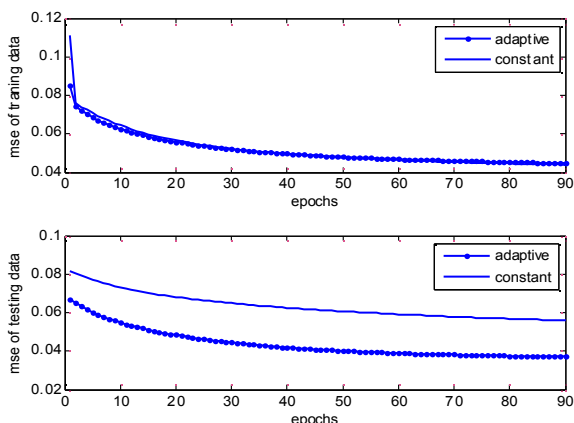
شکل (۴) : میانگین مجموع مربعات خطا در مثال ۱ با رعایت شرایط همگرایی در شبکه عصبی پرسپترون و با در نظر گرفتن باقیمانده بسط سری تیلور بصورت ثابت



شکل (۵) : میانگین مجموع مربعات خطا در مثال ۱ هنگامی که تمام نرخ‌های آموزشی برابر با نرخ آموزش ماکزیمم در محدوده همگرایی باشند.



شکل (۶) : میانگین مجموع مربعات خطا در مثال ۱ در حالتی که نرخ‌های آموزشی از حد مجاز همگرایی بالاتر انتخاب شوند (۱۰۶٪) برابر نرخ آموزش ماکزیمم



شکل (۷) : میانگین مجموع مربعات خطا در مثال ۱ با رعایت شرایط همگرایی و تعیین باقیمانده بسط سری تیلور بصورت تطبیقی و ثابت

شده است. در این شبکه عصبی فرض شده که f^2 تابع خطی و f^1 تابع سیگموئید دو قطبی است. در شکل‌های (۸)، (۹) و (۱۰) مقدار ثابت برای باقیمانده بسط سری تیلور خطای خروجی در نظر گرفته شده است. شکل (۸) همگرایی پارامترهای الگوریتم آموزشی را نشان می‌دهد. در شکل (۹)، نرخ‌های آموزش برابر با حد بالای قیود بدست آمده در بخش ۳ انتخاب شده است.

شکل (۱۰) واگرایی الگوریتم آموزشی در صورت عدم رعایت قیود بدست آمده در بخش ۳ را نشان می‌دهد. تنظیم تطبیقی باقیمانده بسط سری تیلور بر اساس گرادینان نزولی با رعایت شرایط همگرایی در شکل (۱۱) نشان داده شده است.

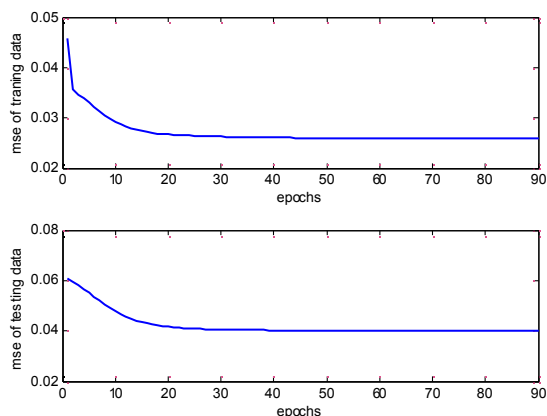
در شکل (۷) همگرایی الگوریتم آموزشی برای باقیمانده بسط سری تیلور بصورت تطبیقی و ثابت نشان داده شده است. همانطور که مشاهده می‌شود با تنظیم تطبیقی باقیمانده بسط سری تیلور، مجموع میانگین مربعات خطا کاهش و سرعت همگرایی افزایش یافته است.

مثال دوم: معادله غیرخطی سیستم بصورت زیر می‌باشد

[۱۵]:

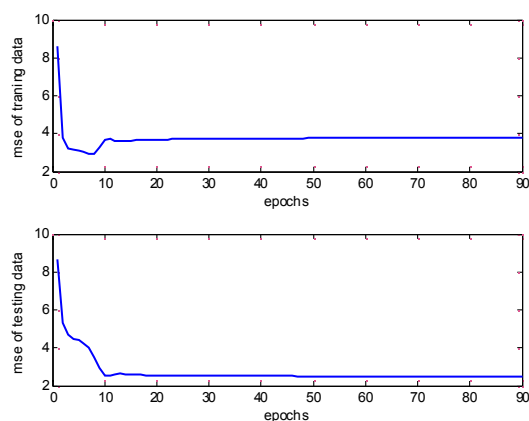
$$y(k+1) = \frac{y(k)}{1+y^2(k)} + u^2(k) \quad (40)$$

که در آن $y(k)$ و $u(k)$ به ترتیب ورودی و خروجی سیستم در زمان k می‌باشند. برای شبیه‌سازی این سیستم غیرخطی از شبکه عصبی با ساختار نشان داده شده در شکل (۲) استفاده شده که در آن n برابر ۲ و p برابر ۱۰ فرض

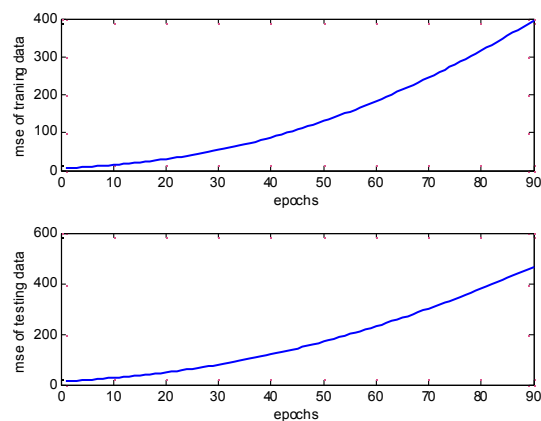


شکل (۸) : میانگین مجموع مربعات خطا در مثال ۲ با رعایت شرایط همگرایی در شبکه عصبی پرسپترون و با در نظر گرفتن باقیمانده بسط سری تیلور بصورت ثابت

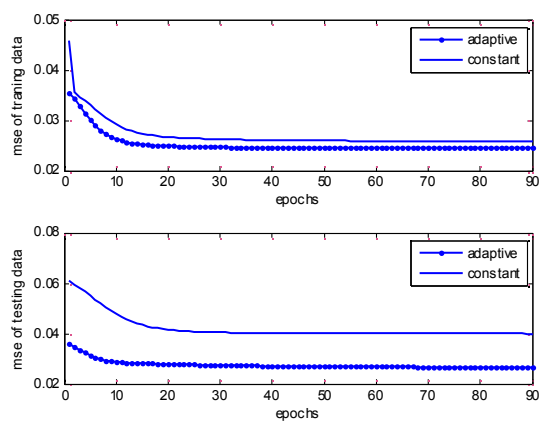
فیلتر غیرخطی تطبیقی عصبی شناساگر با نرخ آموزش بهینه برای همگرایی پارامترها بر پایه گرادیان نزولی



شکل (۹): میانگین مجموع مربعات خطا در مثال ۲ هنگامی که تمام نرخ‌های آموزشی برابر با نرخ آموزش ماکزیمم در محدوده همگرایی باشند.



شکل (۱۰): میانگین مجموع مربعات خطا در مثال ۲ در حالتی که نرخ‌های آموزشی از حد مجاز همگرایی بالاتر انتخاب شوند (۱/۱۰۰٪ برابر نرخ آموزش ماکزیمم)



شکل (۱۱): میانگین مجموع مربعات خطا در مثال ۲ با رعایت شرایط همگرایی و تعیین باقیمانده بسط سری تیلور بصورت تطبیقی و ثابت

- 948, 2009.
- [3] Aliyari Shoorehdeli, M., Teshnehlab, M., Khaki Sedigh, A., "Identification using ANFIS with intelligent hybrid stable learning algorithm approaches", Springer, Neural Comput & Applic 18, pp. 157-174, 2009.
- [4] Bonissone, P.P., chen, Y.T., Goebel, K., khedkar, P.S., "Hybrid Soft Computing System: Industrial And Commercial Application", IEEE, Vol. 87, No. 9, pp.1641-1667, 1999.
- [5] Fine, T.L., Mukherjee, S., "Parameter convergence and learning curves for neural networks", Neural Computation, pp.747-769, 1997.
- [6] Hagan, M.T., Demuth, H.B., Beale, M., "Neural network design", McGraw – Hill publishing company, First edition, 1996.
- [7] Hagan, M.T., Menhaj, M., "Training feedforward networks with Marquardt algorithm", IEEE Transactions on Neural Networks, pp. 989-993, 1994.
- [8] Hykin, S., "Neural Networks: A Comprehensive Foundation", Englewood Cliffs, NJ: Prentice– Hall, 1999.
- [9] Hedjar, R., "Online Adaptive Control of Non-linear Plants Using Neural Networks with Application to Temperature Control System", J. King Saud Univ., Vol. 19, Comp. & Info. Sci., pp. 75-94, 2007.
- [10] Kalluri, S., Arce, G.R., "A General Class of nonlinear Normalized Adaptive Filtering Algorithms", IEEE Trans. Signal Processing, Vol. 47, pp. 2262 – 2272, 1999.
- [11] Kazem, B.I., Zangana, N.F.H., "A Neural Network Based Real Time Controller for Turning Process", Jordan Journal of Mechanical and Industrial Engineering, ISSN 1995-6665, pp. 43 – 55, 2007.
- [12] Kuan, C.M., Hornik, K., "dynamical systems using neural networks", IEEE Transactions on Neural Networks, pp. 4-27, 1991.
- [13] Meireles, M.R.G., Almeida, P.E.M., Simões, M.G., "A Comprehensive Review for Industrial Applicability Of Artificial Neural Networks", IEEE Trans, Vol. 50, No. 3, pp. 585 – 601, 2003.
- [14] Mandic, D.P., "NNGD Algorithm For Neural Adaptive Filters", Electron lett, Vol. 39, No. 6, pp. 845–846, 2000.
- [15] Mandic, D.P., Hanna, A.I. and Razaz Moe, "A Normalized Gradient Descent Algorithm For Nonlinear Adaptive Filters Using A Gradient Adaptive Step Size", IEEE Signal Processing, Letters. Vol. 1, No. 1, pp. 295–297, 2001.
- [16] Mathews, V.J., Xie, Z., "Stochastic Gradient Adaptive filter with Gradient Adaptive Step size", IEEE Transactions. Signal Processing,

در شکل (۱۱) همگرایی الگوریتم آموزشی برای باقیمانده بسط سری تیلور بصورت تطبیقی و ثابت نشان داده شده است. همانطور که مشاهده می شود با تنظیم تطبیقی باقیمانده بسط سری تیلور، مجموع میانگین مربعات خط کاهش و سرعت همگرایی افزایش یافته است.

۵- نتیجه گیری

در این مقاله برای آموزش پارامترهای شبکه عصبی MLP بصورت بهنگام از الگوریتم گرادیان نزولی غیرخطی نرمال تطبیقی بر اساس بسط سری تیلور خطا برای فیلترهای غیرخطی استفاده شده است. در این الگوریتم جمله های مرتبه های دوم و بالاتر بسط سری تیلور خطا در نرخ های یادگیری لایه های فعال ساز میانی و خروجی شبکه عصبی MLP با استفاده از گرادیان نزولی بصورت تطبیقی در گام های مختلف بروز رسانی می شود. مشاهده می شود که این امر دقت و سرعت همگرایی را افزایش می دهد. توسط این الگوریتم محدوده هایی برای نرخ های یادگیری در الگوریتم آموزشی شبکه عصبی MLP حاصل شده است، بطوریکه با انتخاب نرخ یادگیری از این محدوده ها تضمین می شود که الگوریتم آموزشی در فرآیند شناسایی همگرا خواهد بود. نتایج نظری حاصل شده بر روی دو مثال شبیه سازی شده است. نتایج شبیه سازی نشان می دهند که در فرآیند آموزش، دقت و سرعت همگرایی افزایش یافته است. از این روند می توان در کنترل کننده های عصبی بهره برد.

مراجع

- [1] Aliyari Shoorehdeli, M., Teshnehlab, M., Khaki Sedigh, A., Ahmadih Khanesar, M., "Identification using ANFIS with intelligent hybrid stable learning algorithm approaches and stability analysis of training methods", Elsevier, Applied Soft Computing 9, pp. 833–850, 2009.
- [2] Aliyari Shoorehdeli, M., Teshnehlab, M., Khaki Sedigh, A., "Training ANFIS as an identifier with intelligent hybrid stable learning algorithm based on particle swarm optimization and extended Kalman filter ", Elsevier, Fuzzy Sets and Systems 160, pp. 922 –

- Control of a Robot", in Proc. The World Automation Congress - 5th International Symposium on Soft Computing for Industry ISSCI2004, June 28 - July 1, Sevilla, Spain, ISSCI-032 (1:9), 2004.
- [27] Velagic, J., Lacevic, B., Hebibovic, M., "On-Line Identification of a Robot Manipulator Using Neural Network with an Adaptive Learning Rate", in Proc. 16th IFAC World Congress, 03-08 June, Prague, Czech Republic, 2684(1-6), 2005.
- [28] Widrow, B., Lehr, M.A., "Adaptive Neural Networks: Perceptron, Madaline, and Backpropagation", Proceedings of the IEEE, Special Issue on Neural Networks, I: Theory & Modeling; Vol. 78, No. 9, pp. 1415-1442, 1990.
- [29] Wu, W., Feng, G.R., Li, Z.X., Xu, Y.S., "Deterministic Convergence of an Online Gradient Method for BP Neural Networks", IEEE Transaction on Neural Networks; Vol. 16, No. 3, pp. 533-540, 2005.
- [30] Xu, Z., Song, Q., Wang, D., "Recurrent Neural Network Based Tracking Control", Conf. Control, Automation, Robotics and Vision Singapore, 7-10th, 2010.
- [31] Hajiaghasi, S., Paidarnia, H., Abbaszadeh, K., Moatabarian, F., "New method for online interturn faults detection in power transformer with using probabilistic neural network", Computational Intelligence in Electrical Engineering, Vol. 5, No. 2, pp. 1-16, 2014.
- [32] Talebi N, Sadrnia M, Darabi A., "Fault Detection and Isolation of Wind Energy Conversion Systems using Recurrent Neural Networks", Computational Intelligence in Electrical Engineering, Vol. 5, No. 2, pp. 47-68, 2014.
- Vol. 41, pp. 2075 – 2087, 1993.
- [17] Mandic, D., Chambers, J., "Recurrent Neural Networks for Prediction", New York: Wiley, 2001.
- [18] Mpreira, M., Fiesler, E., "Neural Networks with Adaptive Learning Rate and Momentum Terms", Tech. Rep. IDIAP, Vol. 4, 1995.
- [19] Nguyen, V.M.T., "Advanced neural network controllers and classifiers based on sliding mode training algorithms", A thesis, in partial fulfillment of the requirements for the degree of doctor of philosophy from the university of technology, Sydney, 2006.
- [20] Sha, D., Bajic, V.B., "An Optimized Recursive Learning Algorithm For Three-Layer Feedforward Neural Networks For MIMO Nonlinear System Identifications", Intelligent Automation and Soft Computing, Vol. 19, No. 11, pp. 1-15, 2011.
- [21] Song, Q., Wu, Y., Soh, Y. C., "Robust adaptive gradient-descent training algorithm for recurrent neural networks in discrete time domain", IEEE Transactions on Neural Network, pp. 1841–1853, 2008.
- [22] Song, Q., Xiao, J., "On the Convergence Performance of Multilayered NN Tracking Controller", Neural & Parallel Computation, pp.461-470, 1997.
- [23] Sun, X., Liu, Q., Zhang, L., "A BP Neural Network Model Based on Genetic Algorithm for Comprehensive Evaluation", IEEE, 978-1-4577-0856, 2011.
- [24] Torii, M., Hagan, M.T., "Stability of steepest descent with momentum for quadratic functions", IEEE Trans. On Neural Networks, pp. 752-756, 2002.
- [25] Velagic, J., Osmic, N., Lacevic, B., "Neural Network Controller for Mobile Robot Motion Control", World Academy of Science, Engineering and Technology 47, 2008.
- [26] Velagic, J., Hebibovic, M., "Neuro-Fuzzy Architecture for Identification and Tracking

¹ Back propagation (BP)

² Gradient Descent (GD)

³ Real-Time Recurrent Learning

⁴ Robust Adaptive Gradient Descent

⁵ Multivariable Robust Adaptive Gradient Descent

⁶ Normalized Nonlinear Gradient Descent

⁷ Fully Adaptive Normalized Nonlinear Gradient Descent

⁸ Real Time

⁹ Euclidean Norm

¹⁰ Mean Square Error (MSE)