

آموزش شبکه های عصبی مصنوعی به کمک الگوریتم بهینه سازی توده ذرات با سرمایه گذاری (SAPSO)

سید محمدرضا فرشچی
موسسه آموزش عالی سجاد مشهد
انجمن استعداد های ایرانی

Farshchi@iraniantalents.org

مهدی هروی
دانشگاه آزاد اسلامی واحد مشهد
انجمن استعداد های ایرانی

Heravi@iraniantalents.org

مجید بهره پور
باشگاه پژوهشگران جوان
موسسه آموزش عالی خاوران مشهد

Bahrepour@iraniantalents.org

محمد علی میرزا بایابی
پارک علم و فن آوری خراسان
شرکت روبات سازان شرق

Mirzababae@iraniantalents.org

یکی از مدل‌های مرسوم شبکه های عصبی مدل عصبی پیش‌رونده چندلایه¹ می باشد. هنگام کار با شبکه های عصبی پیش‌رونده چندلایه با دو مسئله مواجه می شویم. الف- انتخاب معماری مناسب و ب- انتخاب الگوریتم آموزش مناسب می باشد. معماری به معنی تعداد نرون ها در هر لایه می باشد که انتخاب تعداد نرون ها در هر لایه مبتنی بر مجموعه داده ها² و خصایص آنهاست. روش آموزش متداول در این شبکه ها نیز، پس انتشار خطا می باشد. مشکل الگوریتم پس انتشار خطا همگرایی دیر و توقف در نقاط بهینه محلی می باشد [3]. تا کنون برای رفع مشکل آموزش شبکه های عصبی راهکار های فراوانی ارائه شده است. مثلا الگوریتم های ژنتیک برای تعیین اوزان شبکه های عصبی با یک تکنیک کاهش گرادینان ترکیب شده است و نتایج قابل قبولی داشته است. همچنین الگوریتم جرگه مورچگان³ نیز با همان تکنیک کاهش گرادینان ترکیب شده و نتایج مناسبی داشته است. از الگوریتم جستجوی ممنوعه⁴ برای تعلیم شبکه های عصبی با توپولوژی ثابت استفاده شده است و نتیجه ای بهتر از پس انتشار خطا داشته است. در مقاله ای نیز الگوریتم سرمایه گذاری توده ذرات⁵ با الگوریتم ژنتیک برای تعلیم شبکه های عصبی مقایسه شده است و ثابت شده است که الگوریتم های ژنتیک دارای راندمان بالاتری می باشند [4] در [1] نیز کارولو و همکارانش از PSO استاندارد و PSO با سرعت محدود برای تعلیم شبکه های عصبی استفاده نموده اند، گزارشات اعلام شده برتری این روش را در مقایسه با روش پس انتشار خطا نشان می دهد. انگیزه این مقاله استفاده از ایده جدید هوش جمعی در ترکیب با شبکه های عصبی مصنوعی می باشد تا راهکاری برای غلبه بر چالش موجود در شبکه های عصبی باشد.

چکیده: شبکه های عصبی مصنوعی کاربرد های فراوانی در جداسازی داده ها، انطباق پذیری و یادگیری ماشینی دارند. یکی از چالش های موجود در شبکه های عصبی مسئله آموزش آن می باشد. شبکه می بایست ابتدا آموزش ببیند و سپس بر اساس اندوخته هایش عمل نماید. روش های آموزش آماری همچون پس انتشار خطا (Back Propagation) دارای سرعت پایین و توقف در نقاط بهینه محلی می باشد. از سوی دیگر الگوریتم بهینه سازی توده ذرات با سرمایه گذاری تدریجی (SAPSO) قادر است که با ترکیب دو ایده سرمایه گذاری تدریجی (Simulated Annealing) و بهینه سازی توده ذرات (Particle Swarm Optimization) یک روش غیر قطعی در فضای جواب ها، جوابی نزدیک به بهترین جواب را با سرعت و دقتی بالا پیدا نماید. با توجه به دانش نگارندگان، برای اولین بار است که از SAPSO برای تعلیم شبکه های عصبی استفاده می شود که گزارشات آن در این مقاله ارائه می شود. نتایج شبیه سازی نشان می دهد که این الگوریتم به طور متوسط 30% سریع از روش پس انتشار خطا و به طور متوسط 20% صحت بهتری از روش پس انتشار خطا دارد.

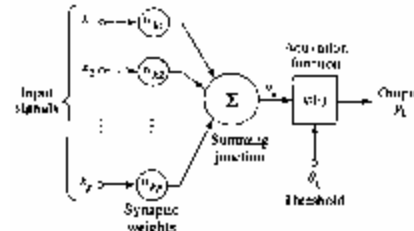
واژه های کلیدی: شبکه های عصبی - بهینه سازی توده ذرات - سرمایه گذاری تدریجی - روش پس انتشار خطا - PSO - SAPSO

۱- مقدمه

شبکه های عصبی مصنوعی دارای ویژگی های فراوانی از جمله انطباق پذیری، قابلیت یادگیری و تعمیم می باشد [1]. در حوزه تطابق الگو ها، شبکه های عصبی مصنوعی قادرند که الگو های جدید را بر اساس تعلیم قبلی خود به کلاس های مرتبط طبقه بندی نمایند.

۲- شبکه های عصبی پیشرونده چند لایه⁶

شبکه های عصبی پیشرونده چندلایه، متشکل از تعداد زیادی اجزای پردازش سیگنال به هم مرتبط به نام نرون ها می باشد [2]. به طور کلی نرون ها ورودی های وزن داده شده خود را با هم جمع می نمایند، حاصل را به یک تابع فعال سازی می دهند و نتیجه را به لایه بعدی منتقل می نمایند. این عمل در شکل 1 نمایش داده شده است.



شکل 1) نمایشی از یک نرون

بهره برد که در این مقاله از مدل ترکیبی بهینه سازی توده ذرات با سرمایه گذاری تدریجی برای آموزش شبکه های عصبی استفاده شده است. این مدل ترکیبی در سال 2004 توسط ونگ و لی ارائه شده است. این محققین کارا بودن روش پیشنهادی خود را بر روی توابع Bench Mark در [9] ارائه نموده اند.

3-1 الگوریتم بهینه سازی توده ذرات

PSO برای اولین بار توسط Kennedy و Eberhart در سال 1995 به عنوان یک روش جستجوی غیر قطعی برای بهینه سازی تابعی مطرح گشت [4]. این الگوریتم از حرکت دسته جمعی پرندگانی که به دنبال غذا می باشند الهام گرفته شده است. در مطالعات بعدی این محققین سعی شد که تئوری و اساس این تعامل گروهی بررسی شود که نتیجه این بررسی نمایشگر این حقیقت بود که تبادل اطلاعات و تجدید نظر در تصمیمات فردی، انطباق با محیط را افزایش می دهد.

```

1: randomly initialize population of particles
2: repeat
3:   for each particle i of the population do
4:     if  $f(x_i(t)) < f(p_i(t))$  then
5:        $p_i(t) = x_i(t)$ 
6:     end if
7:     if  $f(p_i(t)) < f(g(t))$  then
8:        $g(t) = p_i(t)$ 
9:     end if
10:  end for
11:  update velocity and position of each particle according to eqs. (1) and (2)
12: until stop criteria being satisfied
    
```

شکل 2) الگوریتم کلی PSO

روش کار این الگوریتم بدین صورت است. ابتدا تعدادی ذره ایجاد می کنیم و به آنها مقادیر تصادفی می دهیم. در تکرار های بعدی بهترین ذره به سایر ذرات کمک می کند و حرکت آنها را اصلاح می کند و پس از تکرار های متوالی مسئله به جواب بهینه همگرا خواهد شد. الگوریتم کلی PSO در شکل 2 نمایش داده شده است [1].

پارامتر های الگوریتم فوق عبارت است از:

$x_i(t) \in \mathbb{R}^n$ موقعیت فعلی ذره (جوابی که تا کنون بدست آورده است).

$v_i(t) \in \mathbb{R}^n$ سرعت حرکت ذره.

$y_i(t) \in \mathbb{R}^n$ بهترین جوابی که این ذره تا کنون بدست آورده است.

شبکه های عصبی پیشرونده چند لایه یکی از انواع بسیار مهم شبکه های عصبی ایستا محسوب می شوند و کاربرد های فراوانی در مسائلی همچون شناسایی سیستمی،⁷ برابر سازی کانالهای ارتباطی⁸، کنترل و غیره دارند [2]. این مدل از شبکه های عصبی مصنوعی از اوایل 1990 طرح ریزی و بررسی شده است اما هنوز هم مسائلی از این نوع معماری بدون حل باقی مانده است. این شبکه های مدل آماری از تبدیل ورودی ها به خروجی ها می باشد که هر ورودی را به یک خروجی نگاشت می کند و از عناصر حافظه متشکل نشده است. الگوریتم پس انتشار خطا یا BP به عنوان الگوریتم آموزش این شبکه ها به همراه این مدل عرضه شد. هدف از این الگوریتم یک راهکار کاهش گرادین برای حداقل کردن خطا های شبکه می باشد.

۳- هوش جمعی

هوش جمعی خاصیتی است سیستماتیک که در این سیستم، عامل ها به طور محلی با هم همکاری می نمایند و رفتار جمعی تمام عامل ها باعث یک همگرایی در نقطه ای نزدیک به جواب بهینه سراسری می شود [3]. نقطه قوت این الگوریتم عدم نیاز به یک کنترل سراسری می باشد. هر ذره (عامل) در این الگوریتم ها خود مختاری نسبی دارد که می تواند در سراسر فضای جواب ها حرکت کند و می بایست با سایر ذرات (عامل ها) همکاری داشته باشد. دو الگوریتم مشهور هوش جمعی، بهینه سازی جرگه مورچگان⁹ و بهینه سازی توده ذرات¹⁰ می باشند. از هر دو این الگوریتم ها می توان برای تعلیم شبکه های عصبی

ابزار مناسبی جهت محک زدن این روش با روش پس انتشار خطا می باشد که در نتایج نشان داده شده است که برای بهبود پاسخ شبکه در روش بهینه سازی توده ذرات با سرمایه تدریجی کافی است تعداد ذرات را بیشتر کرد ولی در روش پس انتشار خطا می بایست ساختار را عوض نمود.

۱-۴ شبیه سازی های انجام شده

برای شبیه سازی این ایده از Visual Studio 2005 و زبان C# 2.0 استفاده شده است.

در این پیاده سازی، تعلیم شبکه های عصبی به صورت همزمان¹⁴ انجام پذیرفته شده است. یعنی به ازای هر یک ورودی یا یک سطر از مجموعه داده ها تمام اوزان تغییر می نمایند. هر ذره یک ساختار است و آرایه ای از این ساختار، مجموعه ای از تمام ذرات می باشد.

هر ذره مقدار خودش را به کمک معادله (1) و (2) و (3) در هر تکرار اصلاح می نماید.

۵- نتایج شبیه سازی

برای محک بهتر عملکرد روش بهینه سازی توده ذرات با سرمایه تدریجی نتایج این روش، با نتایج روش پس انتشار خطا و همچنین PSO استاندارد مقایسه شده است. سرعت همگرایی روش آموزش شبکه های عصبی به کمک بهینه سازی توده ذرات بسیار بهتر از روش پس انتشار خطا بوده است و روش بهینه سازی توده ذرات با سرمایه تدریجی می باشد ولی دقت الگوریتم بهینه سازی توده ذرات با سرمایه تدریجی بسیار بالاتر می باشد. همچنین در حالت کلی SAPSO از BP سریعتر و دقیقتر است. نتایج بدست آمده در جدول 1 نمایش داده شده است.

روش	روش PSO استاندارد	روش BP	روش SAPSO
زمان آموزش	کمتر از 4 ثانیه برای 530 رکورد و 100 ذره	10 ثانیه برای 530 رکورد	کمتر از 6 ثانیه برای 530 رکورد و 100 ذره
صحت دسته بندی اطلاعات	83% با 100 ذره	72%	88% با 100 ذره آموزش
بهترین جواب بدست آمده با اصلاح ساختار و زمان تعلیم	88% به کمک افزایش ذرات	74% با افزایش تعداد لایه های سطح میانی	92% با افزایش ذرات

$\hat{y}(t) \in \mathbb{R}^n$ بهترین جوابی که سایر ذرات بدست آورده اند.

r_1 and r_2 دو عدد تصادفی بین صفر و یک که با یک تابع توزیع گوسی ایجاد می شوند.

$0 < c_1, c_2 \leq 2$ دو پارامتر که نرخ یادگیری را مشخص می نماید.

w scalar یک مقدار اسکار که از 0.9 شروع شده و تا 0.4 کاهش می یابد. این اسکار باعث می شود که بتوانیم در فضای مسئله گام های بلندتری برداشت و هنگام همگرایی نیز یک جستجو در اطراف نقطه بهینه بتوان داشت.

رفتار هر ذره در هر تکرار تحت تاثیر سایر ذرات قرار می گیرد. یعنی یک همکاری بین تمام ذرات در هر تکرار وجود دارد. هر ذره در هر تکرار با تابع زیر موقعیت خودش را به روز رسانی می کند:

$$v_{ij}(t+1) = w v_{ij}(t) + c_1 r_1 (p_{best}(t) - x_{ij}(t)) + c_2 r_2 (g_{best}(t) - x_{ij}(t)) \quad 1 < i < s, 1 < j < n. \quad (1)$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \quad 1 \leq i \leq s, 1 \leq j \leq n. \quad (2)$$

معادله (1) سرعت حرکت ذره را تعیین می کند و معادله (2) بر اساس سرعت تعیین شده ذره را جابجا می نماید. جابجایی ذره به معنی تغییر مقدار آن ذره است مثلاً در تعلیم شبکه های عصبی وزن هر نرون یک ذره است که آن وزن در تکرار های مختلف تغییر می نماید یا به اصطلاح در فضای جواب جابجا می شود. این جابجایی تا زمانی انجام می پذیرد که به جواب بهینه نزدیک شویم و یا زمان مورد نظر خاتمه یابد.

۴- آموزش شبکه های عصبی مصنوعی به کمک بهینه

سازی توده ذرات با سرمایه تدریجی

برای آموزش شبکه های عصبی به کمک SAPSO می بایست ساختار شبکه عصبی را تعیین و هر یک شبکه را، یک ذره در نظر گرفت. معماری شبکه های عصبی پیشرونده چند لایه بر اساس مجموعه داده های ورودی اش مشخص می شود. در شبیه سازی انجام شده، مجموعه داده های دیابت¹¹ در نظر گرفته شده است. این مجموعه داده¹² دارای 8 مشخصه می باشد و هدف تشخیص وجود و یا عدم وجود بیماری دیابت در افراد می باشد. پس تعداد نرون های لایه ورودی 8 نرون می باشد و در لایه خروجی تنها یک نرون کافی است که مشخص کند فرد به دیابت مبتلا است یا خیر. مطالعات انجام شده نشان دهنده این امر بودند که داشتن یک یا دو نرون در لایه میانی¹³، می تواند بهترین معماری برای این مجموعه داده باشد. در شبیه سازی انجام شده، برای لایه میانی تنها یک نرون در نظر گرفته شد. در نظر گرفتن یک نرون

۷- مراجع

- [1] "An Analysis of PSO Hybrid Algorithms for Feed-Forward Neural Networks Training". Marcio Carvalho, Teresa B. Ludermir. Proceedings of the Ninth Brazilian Symposium on Neural Networks (SBRN'06), 2006 IEEE
- [2] Static and Dynamic Neural Networks from fundamental to advanced theory, Madan M. Gupta, Liang Jin, and Noriyasu Homma, Forwarded by Lotfi A. Zadeh. IEEE Press, 2003
- [3] "Research on Fault Diagnosis of Gearbox Based on Particle Swarm Optimization Algorithm" Pan Hongxia, Ma Qingfeng, Wei Xiuye ICM 2006 * IEEE 3rd International Conference on Mechatronics
- [4] "Adapting Particle Swarm Optimization to Stock Markets" Jovita Nenortaitė, Rimvydas Simutis, Proceedings of the 2005 5th International Conference on Intelligent Systems Design and Applications (ISDA'05) 2005 IEEE
- [5] "Nonlinear System Identification Based on B-Spline Neural Network and Modified Particle Swarm Optimization" Leandro dos Santos Coelho, Renato A. Krohling - International Joint Conference on Neural Networks, Canada July 16-21, 2006 IEEE
- [6] James Kennedy and Russell C Eberhart, Swarm Intelligence, Morgan Kaufman Press, 2001
- [7] Xiaohui Hu, Particle Swarm Optimization Tutorial, Online: <http://web.ics.purdue.edu/~hUx/tutorials.html>, 2002
- [8] Mahamed.H.Omran., Particle Swarm Optimization Methods for Pattern Recognition and Image Processing, Phd Thesis -University of Pretoria ,2004
- [9] XI-HUAI WANG, JUN-JUN LI, HYBRID PARTICLE SWARM OPTIMIZATION WITH SIMULATED ANNEALING, Proceedings of the TIIid International Conference on Machine Learning and Cybernetics, Shanghai, 26-29 August 2004

^۱ Multi Layered Feed forward Neural Networks Dataset

^۲ Ant Colony Optimization
^۳ Tabu Search

^۴ Simulated Annealing

^۵ Multi Layered Feed forward Neural Networks

^۶ System Identification

^۷ Channel Equalization

^۸ Ant Colony Optimization

^۹ Particle Swarm Optimization

^{۱۰} Pima Indian Diabetes Data Set

^{۱۱} Dataset

^{۱۲} Hidden Layer

^{۱۳} Online

Pima Indian Diabetes' Dataset

1-5 بررسی نتایج بروی مجموعه داده های تشخیص خودرو

با تغییر ساختار نرم افزار امکان سنجش برای مجموعه داده تشخیص خودرو نیز فراهم گردید که نتایج این بررسی در جدول 2 بیان شده است.

جدول (2) مقایسه نتایج آموزش شبکه های عصبی برای مجموعه داده های تشخیص خودرو			
روش SPSO	روش PSO استاندارد	روش BP	
کمتر از 90 ثانیه برای 1000 رکورد و 200 ذره و 50 نرون لایه میانی	کمتر از 30 ثانیه برای 1000 رکورد و 200 ذره و 50 نرون لایه میانی	160 ثانیه برای 1000 رکورد با 50 نرون لایه میانی	زمان آموزش
90% با 200 ذره و 50 نرون لایه میانی	78% با 200 ذره و 50 نرون لایه میانی	60% با 50 نرون لایه میانی	صحت دسته بندی اطلاعات
93% به کمک افزایش ذرات	85% به کمک افزایش ذرات	68% با افزایش تعداد لایه های سطح میانی به تعداد 200 نرون	بهترین جواب بدست آمده با اصلاح ساختار و زمان تعلیم
Car Evaluation Dataset			

۶- نتیجه گیری

شبکه های عصبی مصنوعی ابزار های مناسبی جهت انطباق، یادگیری و دسته بندی اطلاعات می باشند. بسیاری از محققین تمایل زیادی در استفاده از این ابزار را دارند اما چالش آموزش شبکه های عصبی را در پیش روی دارند [5]. ترکیب دو ایده هوش جمعی و شبکه های عصبی مصنوعی می تواند پاسخی برای این چالش محسوب شود. نتایج شبیه سازی نمایش می دهد که الگوریتم بهینه سازی توده ذرات می تواند صحت پاسخ شبکه را افزایش و زمان آموزش را کاهش دهد.