http://artificial.ir
اولین سایت هوش مصنوعی

ارائه یک الگوریتم ترکیبی برای پیش‌بینی با استفاده از شبکه‌های عصبی و الگوریتم ژنتیک

مسعود یقینی[۱]؛ شایان کواکب[۲]؛ ناهید غضنفری[۳]

چکیده

گرچه شبکه‌های عصبی مصنوعی یکی از مهمترین تکنیک‌های داده‌کاوی بشمار می‌آیند، با این وجود جستجو برای انتخاب یک ساختار شبکه عصبی مصنوعی بهینه یک موضوع مهم و مورد بحث در زمینه شبکه‌های عصبی مصنوعی می‌باشد. انتخاب ساختار بهینه شبکه‌های عصبی نیاز به فرد خبره و بکارگیری از روش سعی و خطا دارد. تعداد کم نرون‌ها در لایه‌های مخفی منجر به کاهش قدرت شبکه‌های عصبی در یادگیری و در عین حال زیاد بودن تعداد نرون‌ها در لایه‌های مخفی منجر به یادگیری تمرکز یافته بر روی داده‌های آموزشی می‌شود. استفاده از الگوریتم ژنتیک یک ابزار قوی برای یافتن ساختار بهینه شبکه‌های عصبی می‌باشد. در این تحقیق یک الگوریتم ترکیبی شبکه عصبی تکاملی ارائه شده است. در این الگوریتم با استفاده از الگوریتم ژنتیک مقدار بهینه تعداد لایه‌های مخفی، تعداد نرون‌ها در لایه‌های مخفی، نرخ یادگیری، اندازه حرکت و بازه وزن‌های اولیه تعیین می‌شوند زیرا انتخاب بهینه پارامترهای مذکور توانایی جستجو الگوریتم آموزش پس انتشار را افزایش می‌دهند. سپس الگوریتم پیشنهادی با روش ابتکاری انتخاب مدل و تعدادی از الگوریتم‌های معروف در زمینه پیش‌بینی مقایسه شده است.

کلمات کلیدی: داده‌کاوی، پرسپترون چند لایه، الگوریتم آموزش پس انتشار، الگوریتم ژنتیک

Presenting A Hybrid Algorithm for Prediction Using Neural Network and Genetic Algorithm

Masoud Yaghini; Shayan Kavakeb

Abstract

Although *Artificial Neural Networks* (ANNs) are important data mining techniques, the search for the optimal ANN is a challenging task. The ANN should learn the input–output mapping without overfitting the data and training algorithms may get trapped in local minima. Usually the optimum network topology is decided by expert or trial and error approach. The use of too many neurons in hidden layers may overfit the data, which causes the loss of generalization capability of network on the other hand if the number of neuron in hidden layers is not enough then the network may not be able to learn. The use of *Evolutionary Computation* is a promising alternative for ANN topology optimization. Topology parameters of ANN such as the range of initial weights, the learning rates, momentum, the number of hidden layers, the number of neurons in hidden layers can strongly affect the solution searching ability of a *Back Propagation Algorithm*. Optimizing them by genetic algorithm can enhance the probability of finding a globally optimum solution. In this research, we present new *Evolutionary Neural Network Algorithm*. Competitive results were achieved when compared with a heuristic model selection and other *Data Mining Algorithms*.

Key words: Data Mining, Multilayer perceptron, Back Propagation algorithm, Evolutionary algorithms

---

[۱] استادیار دانشکده راه آهن دانشگاه علم و صنعت ایرانYaghini@iust.ac.ir

[۲] دانشجوی کارشناسی ارشد حمل و نقل ریلی دانشگاه علم و صنعت ایران Shayan.Kavakeb@Gmail.com

[۳] دانشجوی کارشناسی ارشد مدیریت اجرایی دانشگاه علم وصنعت ایران n_ghaz@ayhoo.com

۱ مقدمه

شبکه‌های عصبی[۱] در ابتدا به وسیله روانشناسان و عصب شناسان به منظور شناخت و تست محاسبات نرون[۲] مورد بررسی قرار گرفت. به طور کلی شبکه‌های عصبی را می‌توان مجموعه واحدهای ورودی و خروجی متصل به یکدیگر در نظر گرفت که هر اتصال شامل یک وزن می‌باشد. در طول مرحله آموزش با تنظیم مقدار وزن هر اتصال توانایی پیش‌بینی افزایش می‌یابد. شبکه‌های عصبی شامل تعدادی از پارامترهاست که تعیین این پارامترها به صورت تجربی تعیین می‌شوند که می‌توان از این پارامترها به ساختار[۳] شبکه عصبی اشاره نمود (J.Han,M.Kamber, ۲۰۰۶). شبکه‌های عصبی به خاطر امکان ضعیف در تفسیر پذیری مورد انتقاد قرا گرفته‌اند. به عنوان مثال توصیف سمبل‌هایی نظیر وزن‌های سیناپسی و لایه‌های مخفی[۴] در شبکه برای انسان مشکل می‌باشند. این ویژگی‌ها در ابتدا منجر به عدم تمایل استفاده از شبکه‌های عصبی برای داده‌کاوی[۵] شده بود (J.Han,M.Kamber, ۲۰۰۶).

از ویژگی‌های شبکه‌های عصبی می‌توان به قدرت زیاد در مواجه با داده‌های نویزی[۶] اشاره کرد. همچنین زمانی که اطلاعاتی کمی در زمینه ارتباط بین ویژگی‌ها[۷] و دسته‌ها[۸] موجود است، می‌توان از شبکه‌های عصبی به عنوان یک ابزار قدرتمند استفاده نمود (J.Han,M.Kamber, ۲۰۰۶). شبکه‌های عصبی برخلاف درخت‌های تصمیم[۹] بسیار مناسب برای مسائلی با ورودی و خروجی‌هایی از نوع پیوسته می‌باشند. شبکه‌های عصبی در مسایل زیادی با داده‌های واقعی موفق ظاهر شدند (J.Han,M.Kamber, ۲۰۰۶). در سال‌های اخیر تکنیک‌های زیادی به منظور استخراج قوانین از شبکه‌های عصبی آموزش دیده توسعه داده شده است که ضعف اصلی شبکه‌های عصبی را پوشش داده است. با بیان مطالب فوق می‌توان گفت که شبکه‌های عصبی به منظور دسته‌بندی و پیش‌بینی عددی در داده‌کاوی بسیار مناسب می‌باشد.

۲ تعریف مساله

همانطور که اشاره شد روش کلی جهت تعیین پارامترهای شبکه‌های عصبی نظیر ساختار شبکه، نرخ آموزش[۱۰]، اندازه حرکت[۱۱] و بازه وزن‌های اولیه[۱۲] وجود ندارد. در این تحقیق به منظور تعیین بهینه پارامترهای فوق از الگوریتم ژنتیک استفاده شده است. در ابتدا پارامترهای شبکه عصبی به صورت یک کروموزوم کدبندی کرده‌ایم. سپس با استفاده از عملگرهای الگوریتم ژنتیک به سمت پارامترهای بهینه شبکه عصبی حرکت می‌کنیم تا در نهایت به جواب بهینه نزدیک شویم. نتایج حاصل از حل مساله را با برخی از الگوریتم‌های دسته‌بندی با استفاده از تعدادی از مجموعه داده‌های استاندارد مقایسه کرده‌ایم.

۳ مرور ادبیات موضوع

شبکه‌های عصبی مصنوعی مجموعه‌ای ارتباطی[۱۳] است که با الهام از ساختار مغز انسان ساخته شده است. به طور مشخص، پرسپترون چند لایه محبوب‌ترین نوع ساختار شبکه‌های عصبی می‌باشد. به طوریکه نرون‌ها در لایه‌ها قرار گرفته‌اند و فقط اتصال‌های به سمت جلو وجود دارد. علاقه به پرسترون‌های چند لایه با کشف الگوریتم پس انتشار در سال ۱۹۸۶ و سپس پیشنهاد روش‌های بسیاری بر اساس جستجوی گرادیان افزایش یافته است (Miguel Rocha, Paulo Cortez, Joes Neves, ۲۰۰۷). شبکه‌های عصبی مصنوعی به طور گسترده در مسایل مختلف داده‌کاوی بکارگرفته می‌شوند(Han, Morgag, Sinne, ۱۹۹۶) . شبکه‌های پس‌انتشار[۱۴] برای استفاده در داده‌کاوی نسبت به سایر شبکه‌های عصبی بیشتر مورد استفاده قرار می‌گیرند(C.Y.Huang, L.Chen, Y.ChenF.M.chang, ۲۰۰۹) . مشکل اصلی در فرایند آموزش شبکه‌های عصبی احتمال تمرکز بیش از حد بر روی داده‌های آموزشی[۱۵] می‌باشد (M. Chen, Z.Yao, ۲۰۰۸). انتخاب صحیح تعداد نرون‌ها در لایه‌های مخفی و تعداد لایه مخفی از تمرکز بیش از حد بر روی داده‌های آموزشی جلوگیری می‌کند. تمرکز بیش از حد بر روی داده‌های آموزشی زمانی مشاهده می‌شود که درجه آزادی شبکه عصبی بیشتر از آن است که توسط داده‌های آموزشی محدود شود. به عبارت دیگر تمرکز بیش از حد بر روی داده‌های آموزشی زمانی مشاهده می‌شود که روند نزولی در میزان خطا در داده‌های آموزشی و روند صعودی در میزان خطای پیش‌بینی داده‌های تستی و اعتبار سنجی بوجود می‌آید (C.Y.Huang, L.Chen, Y.ChenF.M.chang, ۲۰۰۹). در نتیجه تمرکز بیش از حد بر روی داده‌های آموزشی منجر به ناتوانی در عمومی بودن شبکه‌های عصبی می‌شود (C.Y.Huang, L.Chen, Y.ChenF.M.chang, ۲۰۰۹). شبکه عصبی مصنوعی اطلاعات زیادی از داده‌های آموزشی را در فرایند آموزشی یاد می‌گیرد که این امر منجر به تمرکز بیش از حد بر روی داده‌های آموزشی می‌شود. محاسبه تکاملی[۱۶] یک انتخاب مناسب برای بدست آوردن ساختار بهینه پرسپترون چند لایه می‌باشد که به دلیل قابلیت

جستجوی کلی[۱۷]، جستجوی سریع با کیفیت عالی در فضای جستجو حتی در فضای جستجو بسیار پیچیده می‌باشد. ترکیب محاسبه تکاملی و شبکه‌های عصبی، شبکه‌های عصبی تکاملی نامیده می‌شود (Miguel Rocha, Paulo Cortez, Joes Neves, ۲۰۰۷). الگوریتم ژنتیک به صورت گسترده جهت بهینه سازی شبکه عصبی مورد استفاده قرار می‌گیرد و در بدست آوردن جواب در اکثر مواقع موفق بوده است (C.Y.Huang, L.Chen, Y.ChenF.M.chang, ۲۰۰۹). از الگوریتم ژنتیک به منظور تعیین پارامترها و ساختار شبکه عصبی در جهت افزایش کارایی شبکه عصبی استفاده می‌شود. در بسیاری از مطالعات بهینه‌سازی با استفاده از الگوریتم ژنتیک به همراه الگوریتم پس انتشار و الگوریتم پس انتشار معمولی مقایسه شده است(Arena Caponetto, Fortuna, Xibilia ۱۹۹۲ ، Maniezzo ۱۹۹۴ ، Sexton et al ۱۹۹۸ و Yasuamasa et al ۱۹۹۴). در مطالعات (Gracia et al ۱۹۹۷، M.Rocha et al ۲۰۰۷، Nursel Ozturk ۲۰۰۳، C Hung et al ۲۰۰۹، Jiang et al ۲۰۰۳ و Samanta et al ۲۰۰۳) از الگوریتم ژنتیک به منظور یافتن ساختار بهینه شبکه عصبی استفاده شده است. اکثر این مطالعات بر اساس جستجوی گرادیان جهت یافتن وزن‌های سیناپسی می‌باشد. در مطالعات (S.E.Fahlman, C.Lebiere ۱۹۹۹, M.Frean ۱۹۹۸ , J.siestsma, R.J.F Dow ۱۹۹۸) از روش‌های تخریبی و سازنده جهت طراحی ساختار بهینه استفاده شده است. به طور کلی در روش سازنده با استفاده از شبکه مینیمال (شبکه‌ای با مینیمم تعداد لایه‌های مخفی، گره‌ها و ارتباطات) شروع می‌شود و در زمان لازم لایه مخفی، گره و ارتباطات در طول مرحله آموزش اضافه می‌کنند. در روش تخریبی برعکس فرایند فوق را انجام می‌دهند و با استفاده از یک شبکه ماکزیمال و کاهش لایه‌ها، گره‌ها و ارتباطات ساختار بهین شبکه را پیدا می‌کنند. طراحی شبکه بهینه شبکه عصبی می‌تواند به صورت مسئله جستجو در فضای ساختارهای شبکه عصبی فرمول‌بندی شود به طوریکه هر نقطه بیان‌گر یک ساختار می‌باشد. شرایط بهینه می‌تواند کمترین خطای آموزش، کمترین پیچیدگی شبکه و غیره تعیین شود.(Xin Yao ۱۹۹۹) علل‌های زیر را بیان نموده است که الگوریتم ژنتیک بهترین کاندید برای جستجو در این فضا نسبت به روش‌های تخریبی و سازنده می‌باشد.

- فضا به طور نامحدود بزرگ می‌باشد و تعداد گره‌ها و ارتباطات نا محدود می‌باشد.
- فضا غیر قابل تشخیص[۱۸] می‌باشد به طوریکه تغییرات در گره‌ها و ارتباطات گسسته است و می‌تواند تاثیر غیر پیوسته بر روی شبکه عصبی تکاملی بگذارد.
- فضا بسیار گمراه کننده است به طوریکه ساختارهای مشابه ممکن است نتایج کاملا متفاوت بدهند.
- فضا چند وجهی می‌باشد در نتیجه ساختارهای متفاوت می‌توانند نتایج یکسانی داشته باشند(X.Yao, ۱۹۹۹) .

در این تحقیق از الگوریتم ژنتیک به منظور یافتن ساختار بهینه شبکه عصبی و پارامترهای الگوریتم پس انتشار به طور هم زمان برای دسته‌بندی و پیش‌بینی عددی به کار گرفته شده است. استفاده از این روش که به صورت هم زمان ساختار و پارامترهای الگوریتم پس انتشار مورد تکامل قرار می‌گیرند باعث افزایش قدرت پیش‌بینی شبکه عصبی می‌شود.

۴ الگوریتم تکامل ساختار پرسپترون چند لایه

۴.۱ کدبندی

قبل از استفاده از الگوریتم ژنتیک به منظور بهینه‌سازی پارامترها و ساختار پرسپترون چند لایه بایستی متغیرهای تصمیم تعیین شوند تا فضای جستجوی جواب مشخص شود. در اغلب مسائل پیش‌بینی وجود یک لایه مخفی برای پرسپترون چند لایه به منظور پیش‌بینی کفایت می‌کند. اما در اکثر مسایل پیچیده از دو یا سه لایه مخفی استفاده می‌شود. تعداد نرون‌ها در لایه‌های مخفی وابسته به تعداد لایه‌های مخفی می‌باشد.

در بسیاری از مسایل پیش‌بینی به منظور افزایش توانایی پیش‌بینی پرسپترون چند لایه باید از ساختار بزرگی استفاده شود. در این تحقیق ماکزیمم تعداد لایه‌های مخفی سه لایه در نظر گرفته شده است. پرسپترون چند لایه در این تحقیق به طور کامل مرتبط[۱۹]در نظر گفته شده است. یعنی تمام نرون‌ها در هر لایه به طور کامل با تمام نرون‌ها در لایه بعدی در ارتباط می‌باشند. تابع فعال‌سازی تابع زیگموید در نظر گرفته شده است. ماکزیمم تعداد نرون‌ها در هر لایه مخفی ۶۳ می‌باشد. تعداد نرون‌ها در لایه‌های ورودی و خروجی وابسته به هر مسئله می‌باشد.

چون آموزش پرسپترون‌های چند لایه با بازه‌های مختلف وزن‌های اولیه ولی با ساختار شبکه یکسان منجر به نتایج مختلف می‌شوند، در این تحقیق بازه وزن‌های اولیه مورد تکامل قرار می‌گیرند. زیرا بازه‌های اولیه مختلف این امکان را فراهم می‌سازند که در فرایند آموزش از

مینیمم محلی دور شد و مسئله به سمت مینیمم مطلق هدایت شود. آزمون و خطا معمول‌ترین روش برای تعیین مقدار بهینه نرخ آموزش و اندازه حرکت می‌باشد. اما در این تحقیق مقدار بهینه نرخ آموزش و اندازه حرکت مورد تکامل قرار می‌گیرد و مقدار بهینه آن تعیین می‌شود. در نهایت بازه وزن‌های اولیه، نرخ آموزش، تعداد لایه‌های مخفی، تعداد نرون‌ها در لایه مخفی اول، تعداد نرون‌ها در لایه مخفی دوم و تعداد نرون‌ها در لایه مخفی سوم شبکه عصبی موجود درجدول ۴-۱ به زیر رشته فرم ژن‌گونه نگاشته می‌شود تا احتمال پیدا کردن مینیمم مطلق افزایش پیدا کند.

جدول ۴-۱ بازه پارامترها و تنظیمات ساختار شبکه عصبی

نام پارامترها و ساختار	بازه
بازه وزن‌های اولیه	۰≤ عدد حقیقی <۱.۰
نرخ آموزش	۰< عدد حقیقی <۱.۰
اندازه حرکت	۰≤ عدد حقیقی <۱.۰
تعداد لایه‌های مخفی	۱≤ عدد صحیح ≤۳
تعداد نرون‌ها در لایه مخفی اول	۱≤ عدد صحیح ≤۶۳
تعداد نرون‌ها در لایه مخفی دوم	۱≤ عدد صحیح ≤۶۳ (اگر تعداد لایه‌های مخفی بزرگتر از یک باشد)
تعداد نرون‌ها در لایه مخفی سوم	۱≤ عدد صحیح ≤۶۳ (اگر تعداد لایه‌های مخفی بزرگتر از دو باشد)

در این تحقیق کروموزوم موجود در الگوریتم ژنتیک را به صورت باینری با رشته‌ای از صفر و یک در نظر گرفته شده است. طرح کدبندی استفاده شده به طوری در نظر گرفته می‌شود که هر زیر رشته با تعریف باینری بیان می‌شود که درشکل زیر نشان داده شده است.

۰۱۰۱۰	۱۰۱۰۱۰۱۰	۰۱۰۱۰۱۰۱۰۱	۱۰	۰۱۰۱۰۱	۱۰۱۰۱۰	۱۰۱۰۱۰
۱	۲	۳	۴	۵	۶	۷

زیر رشته ⟵

شکل ۴-۱ کد بیتی پارامترها و ساختار شبکه

هر کروموزوم دارای ۴۵ بیت می‌باشد. که هر زیر رشته از کروموزوم بیان کننده پارامترها و یا ساختار شبکه می‌باشد. به عنوان مثال بازه اولیه وزن‌ها یک عدد حقیقی بین صفر و یک می‌باشد. که با ۷ بیت نمایش داده می‌شود (زیر رشته ۱). هنگام دکد کردن این رشته باینری به عنوان عدد حقیقی، بزرگترین عدد دکد شده برابر $۲^{۷} - ۱$ می‌باشد در نتیجه باید عدد دکد شده بر عدد ۱۲۷ تقسیم شود تا نتیجه حاصله عددی بین صفر و یک باشد. به طور مشابه ۸ بیتی نمایش دهنده نرخ آموزش و ۱۰ بیتی نمایش دهنده اندازه حرکت مقداری بین صفر و یک می‌باشد. تنظیمات ساختار شبکه نیز تعداد لایه‌های مخفی، تعداد نرون‌ها در لایه مخفی اول، تعداد نرون‌ها در لایه مخفی دوم و تعداد نرون‌ها در لایه مخفی سوم به ترتیب با ۲ بیت، ۶ بیت، ۶ بیت و ۶ بیت نمایش داده می‌شود.

۴.۲ معیار اندازه‌گیری دقت پیش‌بینی

از دو ابزار به منظور بیان تعیین دقت هر مدل پرسپترون چندلایه استفاده می‌شود: **درصد مثال‌های دسته‌بندی شده صحیح[۲۰] (PCCE)** که در مسایل دسته‌بندی مورد استفاده قرار می‌گیرد. **ریشه میانگین مربع خطا نرمال شده[۲۱] (NRMSE)** که در مسایل رگرسیون مورد استفاده قرار می‌گیرد. این ابزارهای اندازه‌گیری دقت، در معادله ۴-۱ نمایش داده شده است:

$$\varphi(\mathrm{i}) = \begin{cases} ۱ \text{ if } \; T_i = P_i, \\ ۰ \text{ else}, \end{cases}$$

$$\mathrm{PCCE} = \sum_{i=۱}^{K} \varphi(\mathrm{i})/\mathrm{K} \quad ۱۰۰ \times (\%),$$

$$\mathrm{RMSE} = \sqrt{\sum_{i=۱}^{K} (T_i - P_i)^۲ / K},$$

$$\mathrm{NRMSE} = \frac{\mathrm{RMSE}}{\sum_{\mathrm{i}=۱}^{\mathrm{K}} \mathrm{T_i}/\mathrm{K}} \quad ۱۰۰ \times (\%),$$

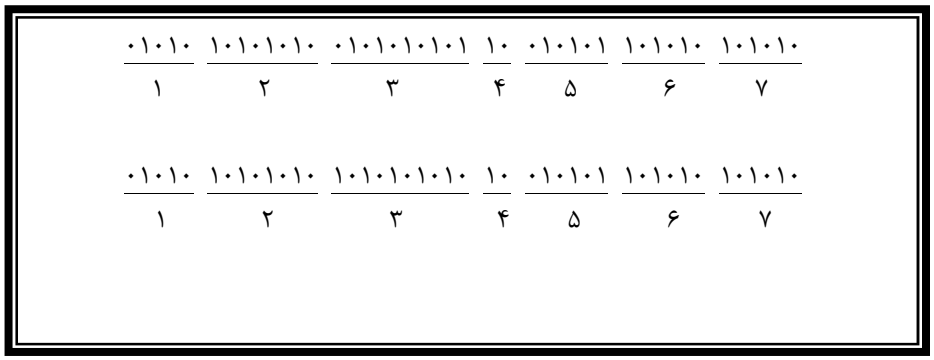
معادله ۴-۱

بطوریکه K تعداد نمونه‌ها، P_{i} و T_{i} مقدار پیش‌بینی شده و واقعی در نمونه i ام می‌باشند. مقدار زیاد PCCE بیان‌گر مدل پیش‌بینی خوب در حالیکه رگرسیون خوب باید مقدار کم NRMSE را داشته باشد. (M.Rocha, P.Cortez, J.Neves, ۲۰۰۴)

۴.۳ الگوریتم پرسپترون چند لایه تکاملی

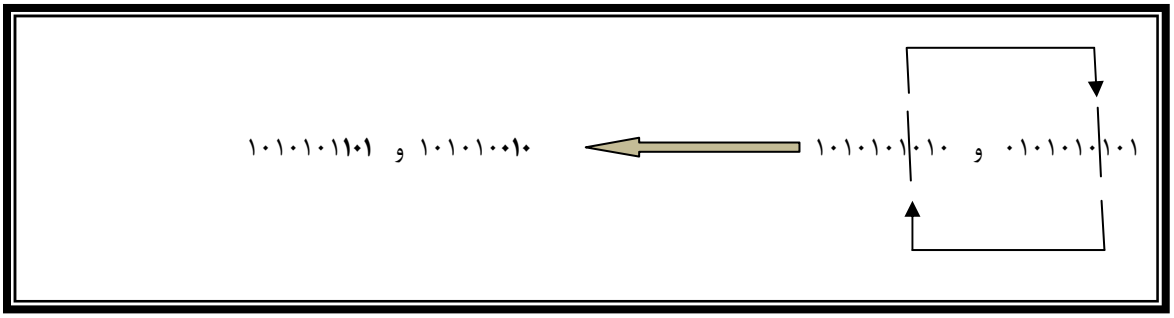
در ابتدا داده‌ها به دو قسمت داده‌های آموزش و داده‌های تستی تقسیم می‌شوند. از مجموعه داده‌های تستی به منظور تعیین میزان خطا استفاده می‌شود. جمعیت اولیه به صورت تصادفی با ۵۰ کروموزوم ساخته می‌شود. سپس هر کروموزوم دکد شده و پرسپترون چند لایه متناظر آن کرومورزوم ساخته می‌شود. هر شبکه عصبی ساخته شده با استفاده از الگوریتم پس انتشار با ۱۰۰۰ تکرار آموزش داده می‌شود. پس از آموزش شبکه‌های عصبی متناظر با کروموزوم‌های جمعیت میزان خطاها با ابزار بیان شده محاسبه می‌شود. با استفاده از روش چرخ رولت[۲۲] به اندازه نصف جمعیت والد انتخاب می‌شود. عملگر جفت‌گیری بر روی هر جفت والدین اعمال می‌شود. عملگر جفت گیری به صورت زیر اعمال می‌شود:

برای هر زیر رشته عملگر جفت‌گیری به صورت جدا اعمال می‌شود. به عبارت دیگر از دو والد دو زیر رشته یکسان به عنوان مثال زیر رشته متناسب با نرخ آموزش انتخاب می‌شود. عملگر جفت‌گیری بر روی این دو زیر رشته اعمال می‌شود و دو زیر رشته جدید ساخته می‌شود. ابتدا یک نقطه تصادفی انتخاب می‌شود که عملگر جفت‌گیری بر روی آن نقطه اعمال شود. جابجایی فقط برروی قسمت اول انجام می‌شود. که در شکل زیر توضیح داده شده است. به عنوان مثال از دو والد زیر دو زیر رشته ۳ انتخاب می‌شود.



شکل ۴-۲ انتخاب دو کروموزوم برای انجام عملگر جفت‌گیری

سپس یک نقطه تصادفی جفت‌گیری انتخاب می‌شود. به عنوان مثال عدد ۳ انتخاب می‌شود. پس نقطه جفت‌گیری تعیین شده است.



شکل ۴-۳ نحوه عمل عملگر جفت‌گیری

همانطور در

شکل ۴-۳ نشان داده شده است پس از تعیین نقطه جفت‌گیری، در این نقطه جفت‌گیری صورت می‌گیرد. احتمال انجام جفت‌گیری ۰.۷ درنظر گرفته شده است. به عبارت دیگر با احتمال ۰.۷ جفت‌گیری انجام می‌شود و با احتمال ۰.۳ جفت گیری انجام نمی‌شود و والدین بدون تغییر باقی می‌مانند.

عملگر دیگری که به کار گرفته می‌شود، عملگر جهش می‌باشد. این عملگر با احتمال ۰.۰۱ جهش را بر روی بیت‌ها انجام می‌دهد. البته جهش را فقط بر روی ۳ بیت اول هر زیر رشته اعمال می‌نماید. این نوع جهش باعث می‌شود تغییر کم‌تری در هر زیر رشته صورت گیرد. تعداد نسل‌ها ۴۰ تا در نظر گرفته شده است.

کد مجازی[۲۳] الگوریتم توضیح داده شده در پاراگراف فوق به شرح زیر می‌باشد:

۱. شروع
۲. تقسیم داده‌ها به دو قسمت داده‌های آموزشی و داده‌های تستی
۳. $I \leftarrow 0$
۴. مقدار دهی اولیه جمعیت $\psi_i \leftarrow$
۵. ارزیابی جمعیت اولیه ψ_i
۶. تازمانیکه(i<G)
۷. انتخاب تعداد P/۲ والد از ψ_i به منظور انجام عملگرهای الگوریتم ژنتیک $\leftarrow A_i$
۸. اعمال عملگرهای جفت‌گیری و جهش بر روی تمام اعضای $A_i \leftarrow O_i$
۹. ارزیابی اولاد در مجموعه O_i
۱۰. انتخاب P/۲-۱ بازمانده ها از $\psi_i \leftarrow S_i$
۱۱. تعیین نسل جدید:
$(\psi_{i+1} \leftarrow \text{best}(\psi_i) \cup S_i \cup O_i)$
۱۲. $i \leftarrow i+1$
۱۳. پایان

ارزیابی برای جمعیت ψ با اندازه P_ψ برابر با :

۱. شروع
۲. $j \leftarrow ۱$
۳. تازمانیکه($j \leq P_{\psi}$)
۴. انتخاب j امین MLP از جمعیت ψ $\longrightarrow MLP_j$
۵. آموزش MLP_j با استاده از داده‌های آموزش با ۱۰۰۰ تکرار
۶. تعیین فیتنس MLP_j با استفاده از RMSE بر روی داده‌های تستی
۷. $j \leftarrow j+۱$
۸. پایان

پرسترون چند لایه تکاملی دسته جمعی با استفاده از ۵۰ تا از بهترین پرسپترون‌های چند لایه تکامل یافته در مراحل تکامل ساخته می‌شود. خروجی شبکه عصبی تکاملی دسته جمعی با استفاده از معادله ۴-۲ ساخته می‌شود.

۴.۴ روش ابتکاری انتخاب مدل

به منظور فراهم کردن پایه‌ای برای مقایسه با الگوریتم پیشنهادی یک روش ابتکاری با رویکرد سعی و خطا ارائه می‌گردد. جاییکه N تا پرسپترون‌های چند لایه مقایسه می‌گردد.

ابتدا پرسپترون چند لایه بهینه با رویکرد سعی و خطا ساخته می‌شود، به عبارت دیگر شبکه پارامترها و ساختار شبکه به صورت سعی و خطا تعیین می‌شوند و پس از آن شبکه عصبی با ۵۰ تا وزن اولیه به طور تصادفی در بازه [۱,۱-] مقدار دهی اولیه می‌شود. سپس با استفاده از الگوریتم پس انتشار به منظور آموزش شبکه‌های عصبی با ۱۰۰۰ تکرار استفاده می‌شود. در نهایت پرسپترون چند لایه با کمترین خطای داده‌های تستی انتخاب می‌شود. از این روش به نام روش ابتکاری شبکه عصبی(HNN)[۲۴] نام برده می‌شود.

سپس از پرسپترون‌های چند لایه آموزش دیده به منظور ساخت شبکه عصبی دسته جمعی ابتکاری[۲۵] (HNNE) استفاده می‌شود. هنگام ساخت شبکه عصبی دسته جمعی دو نکته را باید در نظر گرفت: روش انتخاب استفاده شده برای انتخاب هر مدل و روش ترکیب نتایج هر مدل می‌باشد (Miguel Rocha, Paulo Cortez, Joes Neves, ۲۰۰۷). بدین منظور دو روش زیر در نظر گرفته شده است:

استفاده از شبکه‌های عصبی با ساختار همگن جهت ساخت شبکه عصبی دسته جمعی و آموزش هر شبکه با وزن‌هایی با مقادیر اولیه متفاوت (Miguel Rocha, Paulo Cortez, Joes Neves, ۲۰۰۷)

روش ترکیب خروجی‌های هر مدل با استفاده از میانگین می‌باشد. جاییکه خروجی شبکه عصبی دسته جمعی به صورت زیر می‌باشد:

$$E_{i,j} = (\sum_{k=۰}^{N} S_{i,j,k})/(N+۱)$$

معادله ۴-۲

جاییکه $S_{i,j,k}$ بیان گر j خروجی برای مثال i می باشد. $E_{i,j}$ خروجی نهایی برای مثال i و خروجی گره j می‌باشد. در این تحقیق از این روش به منظور مقایسه الگوریتم پیشنهادی استفاده شده است. در این تحقیق تعداد شبکه‌های عصبی در شبکه عصبی دسته جمعی ۵۰ در نظر گرفته شده است که برابر با تعداد شبکه‌های عصبی در هر جمعیت الگوریتم تکاملی شبکه عصبی می‌باشد. از این روش به منظور ساخت مدل الگوریتم تکامل شبکه عصبی دسته جمعی با استفاده از ۵۰ تا از بهترین شبکه‌های عصبی در طول تکامل استفاده می‌شود.

۴.۵ نتایج

در این تحقیق ۸ مجموعه داده واقعی از منبع UCI که عموما در مسائل داده‌کاوی مورد استفاده قرار می‌گیرند، استفاده شده است. ویژگی‌های هر مجموعه داده آورده شده است.

جدول ۴-۲ ویژگی‌های مجموعه‌های داده

نام مجموعه داده	ورودی			تعداد نمونه‌ها	دسته ها
	Numeric	Binary	Nominal		
Balance(Balance Scale weight and distance)	۴	۰	۰	۶۲۵	۳
Bupa(BUPA liver disoders)	۶	۰	۰	۳۴۵	۲
Car(Car ealuation database)	۰	۰	۶	۱۷۲۸	۴
CMC(Contraceptive method choice)	۵	۳	۱	۱۴۷۳	۳
Abalone(Age of abalone)	۷	۰	۱	۴۱۷۷	رگرسیون
Autos(Auto imports database)	۱۷	۳	۵	۲۰۵	رگرسیون
Heart(Cleveland heart disease database)	۶	۳	۴	۳۰۳	رگرسیون
Mpg(Auto –Mpg(miles per gallon))	۵	۰	۲	۳۹۸	رگرسیون

در دهه اخیر، به دلیل افزایش علاقه به زمینه داده‌کاوی مدل‌ها و الگوریتم‌های زیادی پیشنهاد شده است، هر مدل دارای توانایی‌های متفاوتی بوده و با اهداف مختلفی پیشنهاد شده‌اند. در این تحقیق پنج تا از معروفترین مدل‌ها و الگوریتم‌های داده کاوی انتخاب شده است. از این الگوریتم‌ها به منظور مقایسه با الگوریتم پیشنهادی مورد استفاده قرار گرفته شده است. لیست این الگوریتم‌ها در زیر آورده شده است.

- درخت تصمیم(C۴.۵)
- دستگاه‌های پشتیان بردار
- رگرسیون
- رگرسیون لجستیک

به منظور ساخت مدل‌های فوق از نرم افزار SPSS Clementine ۱۲ استفاده شده است. در هر مدل از تنظیمات Simple استفاده شده است.

روش‌های مذکور با چهار روش زیر که عبارتند از:

- HNN : روش ابتکاری شبکه‌های عصبی
- ENN : روش تکامل شبکه عصبی
- HNNE : روش شبکه عصبی ابتکاری دسته جمعی
- ENNE : روش تکاملی شبکه عصبی دسته جمعی

مقایسه شده‌اند.

نتایج حاصل از اجرای مدل‌های مذکور بر روی هر ۸ مجموعه داده در جداول زیر نشان داده شده است. در جدول ۴-۳ نتایج حاصل از بکارگیری مجموعه داده‌های دسته‌بندی نشان داده شده است. اعداد نمایش داده شده در جدول ۴-۳، PCCE می‌باشد. همانطور که قبلا توضیح داده شد مقدار زیاد PCCE نشان دهنده قدرت بیشتر مدل دسته‌بندی کننده می‌باشد. اعدادی که با شدت رنگ بیشتری نشان داده شده‌اند بیانگر مدل پیش‌بینی کننده بهتر برای مجموعه داده مورد نظر می‌باشد.

جدول ۴-۳ نتایج حاصل از دسته‌بندی

نام مجموعه داده	C۴.۵	Regression logistic	SVM	HNN	ENN	HNNE	ENNE
Balance	۷۸.۳	۸۷.۸	۸۵.۲	۹۴.۹	۹۵.۲	۹۶.۷	**۹۷.۶**
Bupa	۶۶.۵	۶۲.۳	۵۹.۳	۶۸.۵	۶۹.۵	۶۷.۹	**۷۱.۳**
Car	۹۰.۸	۹۱.۲	۹۷.۰	۹۸.۱	۹۷.۸	**۹۸.۹**	۹۷.۹
CMC	**۵۷.۱**	۴۸.۰	۴۹.۵	۵۱.۲	۵۴.۳	۵۵.۶	۵۶.۹
میانگین	۷۳.۱۷۵	۷۲.۳۲۵	۷۲.۷۵	۷۸.۱۷۵	۷۹.۲	۷۹.۷۷۵	**۸۰.۹۲۵**

همانطور که در جدول ۴-۳ مشهود است در دو مورد ENNE بهتر از باقی مدل‌ها نتیجه داده است و در یک مورد HNNE و در یک مورد هم C۴.۵ نتیجه بهتری داده است. در مورد مجموعه داده Balance روش ENNE نسبت به سه روش اول با اختلاف نسبتا زیادی نتیجه بهتری داده است. در مورد مجموعه داده Bupa روش ENNE با اختلاف نسبتا کمتری نتیجه بهتر داده است. در مورد مجموعه داده Car روش HNNE با اختلاف ۰.۱ نسبت به روش ENNE نتیجه بهتری داده است. در مورد مجموعه داده CMC روش C۴.۵ با اختلاف ۰.۲ نسبت به روش ENNE نتیجه بهتری داده است. لازم به ذکر است روش ENN در سه مورد Balance، Bupa، Car نسبت به چهار روش اول نتیجه بهتری داده است. در مورد مجموعه داده CMC روش ENN نسبت به روش های Regression Logestic، SVM و HNN نتیجه بهتری داده است. با توجه به میانگین ارائه شده در سطر آخر جدول ۴-۳ روش ENNE نسبت به سایر روش‌ها مقدار بیشتری دارد.

در جدول ۴-۴ نتایج حاصل از بکارگیری مجموعه داده‌های پیش‌بینی عددی نشان داده شده است. اعداد نمایش داده شده در جدول ۴-۴، NRMSE می‌باشد. همانطور که قبلا توضیح داده شد عدد کوچکتر NRMSE نشان دهنده قدرت بیشتر مدل پیش‌بینی کننده می‌باشد. اعدادی که با شدت رنگ بیشتری نشان داده شده‌اند بیانگر مدل پیش‌بینی کننده بهتر برای مجموعه داده مورد نظر می‌باشد.

جدول ۴-۴ نتایج حاصل از پیش‌بینی عددی

نام مجموعه داده	Regression	SVM	HNN	ENN	HNNE	ENNE
Abalone	۲۲.۴	۲۲.۹	۲۲.۸	**۲۱.۰**	۲۰.۹	۲۱.۰
Autos	۱۹.۴	**۱۲.۵**	۱۳.۲	۱۳.۰	۱۳.۸	۱۲.۷
Heart	۲۱.۰	۲۱.۶	۲۲.۰	۲۱.۸	۲۱.۵	**۲۰.۱**
Mpg	۱۲.۷	۱۳.۶	۱۳.۸	۱۲.۴	۱۲.۰	**۱۱.۷**
میانگین	۱۸.۸۷	۱۷۶۵	۱۷.۹۵	۱۷.۵	۱۷.۵	**۱۶.۳۷**

با توجه به جدول ۴-۴ می‌توان نتیجه زیر را بیان کرد. روش ENNE در سه مجموعه نتیجه بهتری نسبت به سایر روش‌ها نتیجه داده است. در مجموعه داده Abalone دو روش ENNE و ENN نتیجه بهتری داده‌اند. در مجموعه داده Autos روش SVM نتیجه بهتری داده است. در این مجموعه داده روش SVM با اختلاف ۰.۲ نتیجه بهتری داده است. در سه مورد آخر روش ENNE به ترتیب مقادیر ۲۰.۱، ۱۱.۷ و ۱۱.۷ را بدست آورده است که نسبت به باقی روش ها مقدار کمتری را کسب کرده است. در سطر آخر میانگین مقادیر هر روش نشان داده شده است که روش ENNE مقدار بهتری را کسب کرده است.

در نهایت با توجه به نتایج ارائه شده در جدول‌های فوق می‌توان نتیجه‌گیری نمود که روش ENNE نسبت به سایر روش‌ها دقت بیشتری در دسته‌بندی و پیش‌بینی را دارا می‌باشد.

۵ نتیجه‌گیری

در این تحقیق به منظور افزایش دقت پیش‌بینی پرسپترون چند لایه یک روش تکاملی برای تعیین پارامترها و ساختار پرسپترون چند لایه ارائه شد. سپس این الگوریتم پیشنهادی با سایر الگوریتم‌های پیش‌بینی نظیر درخت‌های تصمیم، SVM، رگرسیون و رگرسیون لجستیک

مقایسه شد و همچنین یک مدل ابتکاری به منظور انتخاب مدل پرسپترون چند لایه جهت تعیین معیاری برای مقایسه با الگوریتم پیشنهادی که نتیجه مقایسات بدین گونه بود که نشان داد الگوریتم پیشنهادی و همچنین مدل دسته جمعی نسبت به سایر الگوریتم‌ها از دقت خوبی برخوردار می‌باشد. با ارائه این الگوریتم دیگر نیازی به تعیین پارامتر و ساختار پرسپترون چند لایه به رویکرد آزمون وخطا نمی‌باشد.

۶ پیشنهادات آتی

موارد زیر را می توان به عنوان افق تحقیقات آتی در زمینه الگوریتم شبکه عصبی تکاملی در نظر گرفت:

ساخت مدل دسته جمعی با استفاده از فیتنس هر پرسپترون چند لایه که این امر منجر به افزایش دقت پیش‌بینی می‌شود.

استفاده از سایر روش‌های آموزش پرسپترون چند لایه نظیر RPROP و مقایسه این روش‌ها با یکدیگر به منظور یافتن الگوریتم مناسب‌تر

ارائه تکنیک مکمل به منظوری جلوگیری از افتادن الگوریتم پس انتشار در مینیمم محلی

منابع و مراجع

[۱]C.Soares. (۲۰۰۳). "Is the UCI repository useful for data mining?" *Progress in Artficial Intelligence* , ۲۹۰۲, ۲۰۹-۲۲۳.

[۲]C.Y.Huang, L.Chen, Y.ChenF.M.chang. (۲۰۰۹). "Evaluation the process of a genetic algorithm to improve the back-propagation network: A Mont Carlo study." *Expert Systems with Applications* , ۱۴۵۹-۱۴۶۵.

[۳]D.J.Hand. (۱۹۹۸). "Data mining: Statistics and More?" *The American Statistician* , ۵۲.

[۴]E.Keogh, S.Lonardi, C.A.Ratanamahatana. (۲۰۰۵). "Towards Parameter-Free Data Mining." *Proceeding of the tenth ACM SIGKDD international conference on Knowlledg discobery and data mining* .

[۵]G. Eiben, Jim Smith. (۲۰۰۳). "Introduction to Evolutionary Computing." Springer.

[۶]Han, Morgag, Sinne. (۱۹۹۶). "Optimization of feedforward neural network." *Engineering Application of artificial Intelligence* , ۹, ۱۰۹-۱۱۹.

[۷]I. Tsoulos, Dimitris Gavrilis, E. Glavas. (۲۰۰۸). "Neural Network Construction and Training Using Grammatical Evolution." *Neurocomputing* , ۲۶۹-۲۷۷.

[۸]I.H.Witten, E.Frank. (۲۰۰۵). "Data Mining: Practical MachhinLearning Tools and Technics." Morgan Kaufmann.

[۸]J.D.Schaffer, D.Whitley, L.J.Eshelman. (۱۹۹۲). "Combinations of Genetic Algorithms and Neural Networks: A Survey of State of the Art." *IEEE international joint conference on neural networks* , ۶۶۷-۶۷۳.

[۹]J.Han,M.Kamber. (۲۰۰۶). "Data Mining: Consepts and Technques." Morgan Kaufmann.

[۱۰]J.W.Seifert. (۲۰۰۴). "Data Minig : An Overview." *CRS Report RL۳۱۷۹۸* .

[۱۱]M. Chen, Z.Yao. (۲۰۰۸). "Classification Techniques of Neural Networks Using Improved Genetic Algorithm." *Second International on genetic and EvolutioanryComputing* .

[۱۲]M.Rocha, P.Cortez, J.Neves. (۲۰۰۴). "Ensemble of Artificial Networks with heterogeneous Topologies." *Proceeding of the Fourth Symposium on Engineering of Intelligence Systems* .

[۱۳]M.Schumacher, R.Robner, W.Vach. (۱۹۹۶). "Neural networks and logistic regression." *Computational Statistics & Data Analysis* , ۲۱, ۶۶۱-۶۸۲.

[۱۴]Miguel Rocha, Paulo Cortez, Joes Neves. (۲۰۰۷). "Evolution of neural networks for classification and regression." ۷۰, ۲۸۰۹-۲۸۱۶.

[۱۵]Negnevitsky, M. (۲۰۰۴). "Artificial Intelligence A Guid to Intelligence System". Addison Wesley.

[۱۶]P.A.Catillo, J.J Merelo, M.G.Arenas, G. Romero. (۲۰۰۸). "Comperaring evolutionary Hybrid systems for design and optimization of multilayer perceptron structure along training parameters". *Information Science* .

[۱۷]Reidmiller, M. (۱۹۹۴). "Advanced SupervisedLearning in Multi-Layer Perceptrons From BachPropagation to Adaptive Learning Algorithms." *Int. J. Comput* , ۲۶۵-۲۷۸.

[۱۸]S.Mitra, S.Pal, P.Mitra. (۲۰۰۲). "Data mining in Soft Computing framwork: a survay." *IEEE Trans. Neural Networks , ۱*, ۳-۱۴.

[۱۹]T. Back, D.B Fogel, Z. Michalewicz. (۲۰۰۰). "Evolutionary Computation ۲ Advanced Algorithms and Poerators". Institute of physics publishing.

[۲۰]X.Yao. (۱۹۹۹). "Evolving Artificial Neural Network." *Proceeding of the IEEE* , ۱۴۲۳-۱۴۴۷.

[۲۱]X.Yao, Y.Liu. (۱۹۹۷). "A new evolutioanry Systems for evolving artificial neural network." *IEEE Trans. Neural Networks* , ۶۹۴-۷۱۳.

زیر نویس‌ها

[۱] Neural Networks
[۲] Neron
[۳] Topology
[۴] Hidden Layers
[۵] Data Mining
[۶] Noisy Data
[۷] Attributes
[۸] Classes
[۹] Decision Trees
[۱۰] Learning Rate
[۱۱] Momentum
[۱۲] Range of Initial Weights
[۱۳] Connectionist
[۱۴] Back-Propagation
[۱۵] Overfitting
[۱۶] Evolutionary Computing
[۱۷] Global
[۱۸] NonDifferentiable
[۱۹] Fully Connected
[۲۰] Percentage of Correctly Classified Examples
[۲۱] Normalized Root Mean Squared Error
[۲۲] Roulette Wheel
[۲۳] Pseudo Code
[۲۴] Heuristic Neural Network
[۲۵] Heuristic Neural Network Ensemble