

طبقه‌بندی خودکار طیف ستارگان پایگاه SDSS-DR9 با استفاده از

شبکه عصبی مصنوعی

شکوفه خیردستان^۱، مهدی بازرگان^۱، سهند شکری نیری^۲، فریبا شویکلو^۱

^۱ گروه فیزیک، دانشکده علوم، دانشگاه زنجان

^۲ گروه IT، دانشگاه صوفی رازی زنجان

چکیده

سیستم های هوشمند برای پردازش خودکار داده های حجیم نجوم از جمله طبقه بندی اجسام و محاسبه پارامترهای آنها به طور گسترده استفاده می شوند. داده های وسیع طیف نگاری اجرام، دانشمندان را به استفاده از تکنیک های خودکار برای بررسی این مجموعه داده ها تشویق کرده است. SDSS^۱ یکی از منابع مهم قابل دسترس عمومی به این داده های حجیم است. در این مقاله از شبکه عصبی احتمالی^۲ برای طبقه بندی خودکار ۵۰۰۰ طیف ستاره ای از داده های پایگاه SDSS-DR9 برای بدست آوردن رده و زیر رده آن استفاده کرده ایم. نتایج طبقه بندی خودکار داده ها به صورت کاتالوگ برای استفاده در دسترس عموم ارایه می گردد.

شبکه عصبی مصنوعی به عنوان یک سیستم خودکار امروزه به طور وسیع در شاخه های مختلف نجوم و اختر فیزیک مورد استفاده قرار می گیرند. تخمین انتقال به سرخ [۱]، طبقه بندی طیف ستارگان [۲]، طبقه بندی مورفولوژی کهکشان ها [۳] به عنوان نمونه هایی از کارهای انجام شده بیانگر استقبال گسترده منجمین از این روش ها می باشد.

پیش پردازش داده ها:

داده ها از پایگاه SDSS-DR9 گرفته شده است، که تعداد ۵۰۰۰ طیف از DR9 برای استفاده آموزش شبکه عصبی و تعداد ۵۰۰۰ برای آزمون مورد بررسی قرار می گیرد. داده ها از نوع Fits هستند که یکی از رایج ترین فرمت ها برای تبدیل و ذخیره داده های نجومی است. Fits برای داده های علمی طراحی شده است که ترکیبی از داده های دوتایی و ASCII می باشند. داده های Fits شامل طیف، جداول اطلاعات و تصویر اجرام می باشند که به منظور استفاده در شبکه عصبی، تصاویر طیف ستاره ای باید به ASCII (۲ ستون طول موج و شدت) تبدیل گردد. پس از تبدیل داده ها به ASCII، بازه مشترک طول موج را برای طیف ستاره ها بدست آورده و داده های طیف همه ستارگان را در این بازه مشترک قرار می دهیم. بعد از یکسان سازی بازه طول موج داده های آزمون و آموزش همه ستارگان، تعداد داده های نقطه ای آنها یکسان شده و سپس بهنجارش به واحد روی تمام داده ها به منظور آماده سازی برای ورود به شبکه عصبی صورت می گیرد.

شبکه عصبی احتمالی:

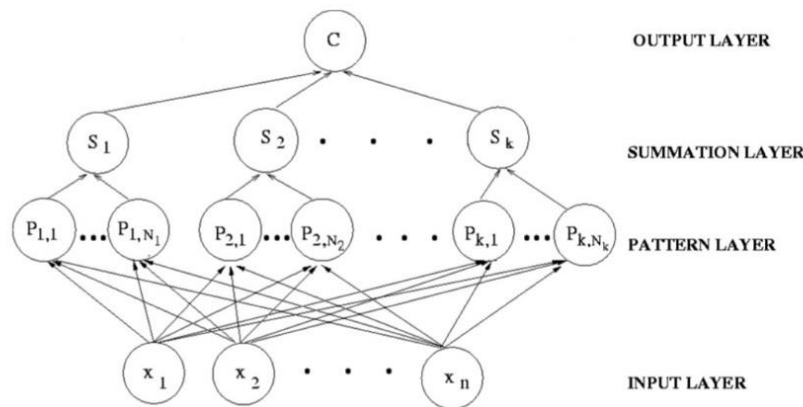
شبکه عصبی احتمالی توسط Specht در سال ۱۹۹۰ توسعه یافت [۴]. این شبکه با پیروی از یک الگوی آماری که طبقه بندی کننده ی بیزین^۳ نام دارد، راه حل جامعی برای مشکلات طبقه بندی فراهم کرده است. ساختار این الگوریتم بر اساس تابع چگالی احتمال است که در حالت کلی شامل چهار لایه است (شکل ۱): یک لایه ی ورودی و سه لایه ی پردازش اطلاعات لایه ی الگو، لایه ی مجموع و لایه ی خروجی. نرونهای لایه ورودی صرفاً وظیفه انتقال مقادیر ورودی به تمام نرونهای لایه ی دوم را بر عهده داشته و هیچگونه پردازشی در این لایه بر روی داده ها صورت نمی گیرد. لایه الگو مجموعه آموزش را سازماندهی می کند به طوری که با استفاده از خروجی لایه ی اول برداری از احتمال ها را به عنوان خروجی شبکه تولید می کند. در لایه ی سوم که لایه ی مجموع می باشد، هر جز با یکی از اجزا پردازش لایه ی الگو

^۱Sloan Digital Sky Survey

^۲ Probabilistic Neural Network

^۳ Bayesian

ترکیب و به کلاس مشابه مربوط می‌شود و مقادیر احتمالات را می‌دهد. سپس بزرگترین عدد احتمال لایه‌ی مجموع در لایه‌ی خروجی ظاهر می‌شود که با یکی از رده‌های طیف آموزش مطابقت می‌کند.



شکل ۱: شبکه عصبی احتمالی

طبقه‌بندی و کدگذاری طیف ستارگان:

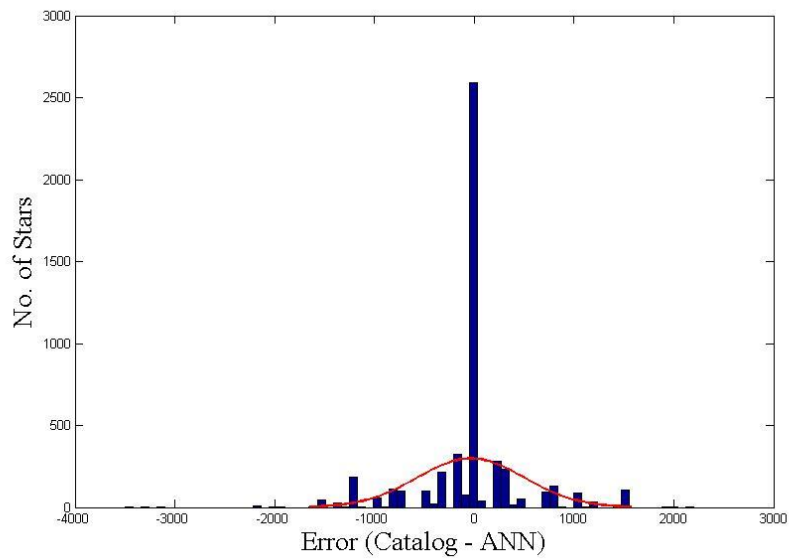
در علم ستاره‌شناسی بر اساس تخمینی که از نور و دمای سطحی ستارگان با استفاده از طیف سنجی نجومی بدست می‌آید، رده بندی ستارگان صورت گرفته است و انواع متفاوتی دارد. ترتیب تقسیم بندی از گرمترین به سردترین O, B, A, F, G, K, M می‌باشد. هر رده خود به زیر رده‌های بیشتری تقسیم می‌شود که این کار معمولاً با نسبت دادن اعداد بین ۰ تا ۹ صورت می‌پذیرد (عدد ۰ برای گرمترین و ۹ برای سردترین). رده و زیر رده ستارگان به ترتیب به صورت حروف لاتین و عدد است که باید به یک کد تبدیل شوند. برای این کار از رابطه زیر استفاده می‌کنیم:

$$CodeNumber = 1000.0 \times A1 + 100.0 \times A2 \quad (1)$$

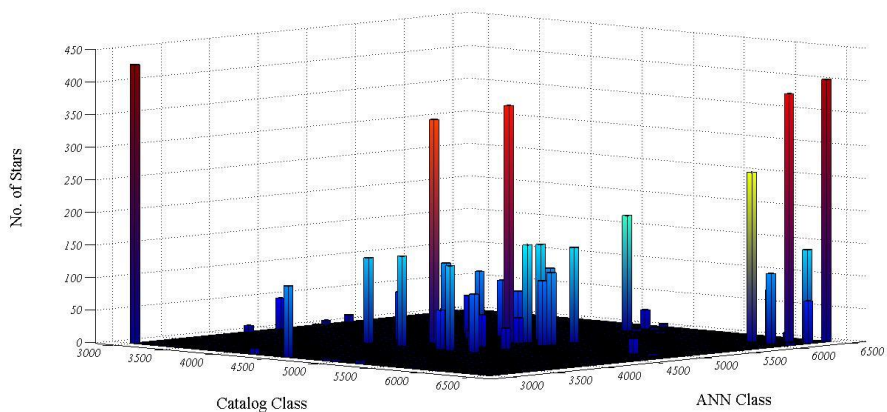
که در آن $A1$ نشان دهنده‌ی نوع ستاره است (O تا M که شماره گذاری آن از ۱ تا ۷ می‌باشد) و $A2$ زیر رده‌ی آنها است (که مقدار آن از ۰ تا ۹ می‌باشد). برای مثال: $F3$ با کد ۴۳۰۰ و $K9$ با کد ۶۹۰۰ نشان داده می‌شوند.

نتیجه‌گیری:

در این کار ما توانستیم تعداد ۵۰۰۰ از طیف ستارگان مجموعه آزمون پایگاه داده SDSS-DR9 را طبقه‌بندی نموده و هر کدام از آنها را به یکی از داده‌های مجموعه آموزش دسته‌بندی کنیم. سپس خطای طبقه بندی شبکه عصبی را با توجه به مشخص بودن رده و زیر رده طیفی هر کدام از داده‌های آزمون و آموزش محاسبه نمودیم که خلاصه این نتیجه دسته‌بندی در شکل (۲) آورده شده است که یک نمودار ستونی خطا بر حسب تعداد ستارگان می‌باشد. همانطور که در این شکل می‌بینیم بزرگترین قله برای خطای صفر می‌باشد و تعداد ۲۵۹۲ از ۵۰۰۰ ستاره طبقه‌بندی درستی را هم برای رده و هم زیر رده طیفی داشته است. با بررسی دقیق‌تر خطاهای طبقه‌بندی که بین بازه $+1150$ و -1150 می‌باشند، حدود یک رده طیفی و نیم زیر رده طیفی خطا را نشان می‌دهد. شکل ۳ خروجی شبکه عصبی را به صورت نمودار پراکندگی سه بعدی نمایش می‌دهد.



شکل ۲: نمودار هیستوگرام طبقه‌بندی ۵۰۰۰ طیف از منبع *SDSS - DR9*



شکل ۳: نمودار پراکنندگی طبقه‌بندی *SDSS - DR9*

مرجع‌ها

1. O'Mill, Ana Laura., Duplancic, Fernanda., García Lambas, Diego., Sodr , Laerte, Jr., *Monthly Notices of the Royal Astronomical*, **2010**, P1395-P1408.
2. Bazarghan Mahdi, *Astrophysics and Space Science*, **2007**.
3. Yanxia, Zhang., Lili, Li., Yongheng, Zhao., *Astronomical Society of India*, **2008**, P233-P239.
4. DONALD F. SPECHT, *Neural Networks*, **1990**.