



کاربرد ترکیب طبقه بندها در طبقه بندی مسائل پیچیده^۱

رضا ابراهیم پور

مدرس گروه الکترونیک دانشگاه

شهید رجایی ودانشجوی دکتری علوم

اعصاب شناختی گرایش رایانش و هوش

مصنوعی، پژوهشکده علوم شناختی،

مرکز تحقیقات فیزیک نظری و ریاضیات

ebrahimpour@ipm.ir

نیما حاتمی

دانشجوی کارشناسی الکترونیک،

دانشگاه شهید رجایی

nimaht62@gmail.com

چکیده- برای افزایش دقت طبقه بندی مسائل پیچیده، می توان از ترکیب طبقه بندهایی استفاده کرد که از یک الگوریتم یادگیری، ولی با پیچیدگی ها و پارامترهای متفاوت بهره می گیرند. در این مقاله از مجموعه مجموعه داده ^۲های ELENA، مجموعه داده تصاویر ماهواره^۳ استفاده شده است که به دلیل بالا بودن بعد بردار ویژگی ها، کاهش افزونگی اطلاعات^۴ توسط روش تحلیل مولفه های اصلی^۵ صورت گرفته است. از مجموعه داده تصاویر ماهواره، بردارهای ویژگی، با ۸۲۰۰ مولفه استخراج می شود و سپس به روش تحلیل مولفه های اصلی، ۳۶ مولفه اول آن انتخاب شده و به سه شبکه عصبی پرسپترون با تعداد نرونهای متفاوت در لایه مخفی و وزنهای اولیه متفاوت اعمال شده و طبقه بندی اولیه صورت می گیرد. در مرحله بعد، نتایج طبقه بندی این سه طبقه بند، به یک شبکه عصبی پرسپترون با یک لایه مخفی به عنوان ترکیب کننده اعمال می شود. نرخ شناسایی شبکه های عصبی پایه ۸۶٪/۸۳٪، ۸۸٪/۸۶٪، ۸۶٪/۸۶٪ و برای سیستم ترکیبی آنها ۸۸٪/۸۸٪ است.

واژه های کلیدی- طبقه بندی، ترکیب طبقه بند ها، مسائل پیچیده، شبکه های عصبی

^۱ Complex Problems

^۲ Data Base

^۳ Satimage data base

^۴ Dimantion reduction

^۵ Principal component analyse(PCA)

۱-مقدمه

یکی از موضوعات مورد علاقه در یادگیری ماشین^۶، چگونگی ترکیب نتایج چند طبقه بند پایه، به منظور رسیدن به دقت طبقه بندی بالا می باشد [۱]. در شناسایی الگو، ترکیب نتایج چندین طبقه بند مکمل یکدیگر، باعث بهبود کارایی طبقه بندی، افزایش نرخ طبقه بندی و افزایش قابلیت اعتماد سیستم تشخیص الگو می شود [۲]. ترکیب طبقه بندها، یکی از روش های افزایش کارایی در زمینه طبقه بندی مسائل پیچیده است. پیچیدگی در مسائل طبقه بندی، می تواند ناشی از بالا بودن بعد ویژگی ها، محدود بودن تعداد نمونه ها، همپوشانی^۷ کلاسها و وجود نویز قابل ملاحظه در نمونه ها باشد.

از مهمترین نیازها در ترکیب، این است که تعدادی طبقه بند با کارایی قابل قبول، همچنین مستقل در تصمیم گیری موجود باشد. به عبارت دیگر، ما به طبقه بندهایی با اطلاعات مختلف و بدون اطلاعات زائد^۸ نیازمندیم. در غیر این صورت طبقه بند ها نتایج یکسانی برای تمامی حالات خواهند داد که این به هیچ وجه برای ترکیب مفید نیست. طبقه بند ها بایستی برای الگو هایی که درست طبقه بندی شده اند نتایج یکسان و برای موارد اشتباه نتایج متفاوت داشته باشند. در مواردی که الگوها توسط همه طبقه بندها به صورت اشتباه طبقه بندی شده اند، بدست آوردن نتیجه صحیح بوسیله سیستم ترکیب امکان پذیر نیست. بخشی از خطاهای قابل حذف در سیستم ترکیبی، نبایستی برای تک تک طبقه بندها مشترک باشد. به همین دلیل کاهش همبستگی^۹ بین طبقه بندها بسیار مهم است.

شبکه های عصبی به عنوان طبقه بند، توسط یک پروسه یادگیری ساخته می شوند. به همین خاطر به منظور داشتن طبقه بندهای متفاوت، بایستی روش یادگیری یا پارامترهای

یادگیری طبقه بندها متفاوت باشد. پروسه یادگیری میتواند تحت تاثیر طریقه نمایش^{۱۰} الگوهای ورودی، نمونه های آموزشی^{۱۱}، مکانیزم یادگیری و استراتژی نظارت بر یادگیری قرار گیرد. در این مقاله تاکید بر مکانیزم های مختلف یادگیری است. تغییرات پروسه یادگیری می تواند توسط موارد زیر انجام شود:

الف- استفاده از الگوریتم های یادگیری متفاوت: برای مثال می توانیم از شبکه عصبی پرسپترون چند لایه، شبکه و درخت تصمیم در کنار هم در یک سیستم ترکیبی استفاده نماییم [۳].

ب- استفاده از یک الگوریتم یادگیری با پیچیدگیهای متفاوت: استفاده از شبکه عصبی با تعداد نرون ها و یا لایه های متفاوت، یا استفاده از تعداد نقاط همسایه گوی متفاوت در طبقه بندهمسانه نزدیکتر، می تواند منجر به ایجاد طبقه بند های متفاوتی بشود.

ج- استفاده از یک الگوریتم یادگیری با پارامترهای مختلف: استفاده از الگوریتم یادگیری تندترین شیب با وزن های اولیه، تعداد دفعات^{۱۲} آموزش، نرخ یادگیری^{۱۳} و ممنوم یا تابع هزینه متفاوت میتواند طبقه بندهای متفاوتی را ایجاد کند [۴].

در این مقاله از یک شبکه عصبی پرسپترون سه لایه با الگوریتم آموزشی تندترین شیب جهت تعیین وزنها طبقه بند های پایه استفاده شده است. پس از آموزش طبقه بند های پایه، یک شبکه عصبی دیگر در پروسه آموزش، مقدار وزنها بهینه برای ترکیب نتایج طبقه بند های پایه را هوشمندانه تعیین می نماید.

بقیه ساختار مقاله به صورت زیر است: در بخش دوم، کاهش افزونگی بررسی می شود. بخش سوم به معرفی داده تصاویر ماهواره اختصاص دارد. بخش چهارم، الگوریتم ترکیب طبقه بند ها را بیان می کند. بخش پنجم به نتایج تجربی و بخش

¹⁰ Representation

¹¹ Training Samples

¹² Epochs

¹³ Learning Rate

⁶ Machine learning

⁷ Overlapped

⁸ Redundant

⁹ Correlation Reduction

ششم به نتیجه گیری اختصاص دارد.

روش PCA ، M بردار ویژه مهم برای دیتارابدست آورد.

این بردارهای ویژه از اولین بردارهای ویژه یافته شده در یک ماتریس $M \times N$ هستند که ترکیب خطی از ویژگیها را در بر می گیرند:

$$L = Y^T Y, L_{mn} = X_m^T X_n \quad (5)$$

M بردار ویژه V_{lk} برای L پیدا می شود. مشخصه های ویژه از رابطه (6) محاسبه می شود.

$$\hat{U}_K = \sum_{k=1}^M V_{lk} X_k \quad l=1, \dots, M \quad (6)$$

این مشخصه ها بر حسب مقادیر ویژه اختصاص داده شده به آنها نرمالیزه و مرتب می شوند. مشخصه ها بوسیله بردار $W_k = U_k^T (T - A)$ به فضای ویژگی جدید تصویر می شوند. سپس این بردار، برای طبقه بندی به ورودی شبکه عصبی اعمال می شود. با آزمایش روی نمونه های مجموعه داده مشخص شد که ۳۶ مولفه اول PCA اطلاعات لازم جهت مسئله طبقه بندی رادر بر دارند.

۳- مشخصات دیتای تصاویر ماهواره

این مجموعه داده یکی از معروف ترین مجموعه داده های پیچیده در ترکیب طبقه ها می باشد. دیتای اصلی برای این مجموعه داده از مرکز کنترل از راه دور استرالیای ناسا، به منظور استفاده در تحقیقات بدست آمده است. این مجموعه داده از تصاویر ماهواره اسکنر چند طیفی^{۱۴}، که یکی از منابع مهم اطلاعات است، استخراج شده است. این مجموعه داده برای مقایسه عملکرد ماشین یادگیری، الگوریتم های استاتیکی و شبکه های عصبی در حوزه های صنعتی شامل پزشکی، تجاری، پردازش تصویر و طراحی مهندسی مورد

۲- کاهش افزونگی

یکی از مهمترین روش ها برای کاهش افزونگی جهت تسهیل پردازش داده و طبقه بندی، استفاده از تحلیل مولفه های اصلی، PCA است. این تبدیل به طور گسترده در آنالیز داده، جهت کاهش ابعاد استفاده می شود و یک تبدیل بهینه به مفهوم بسته بندی اطلاعات می باشد. تبدیل در عین حال که ابعاد داده را کاهش می دهد، عمده اطلاعات داده اصلی را نیز در بر دارد.

فرض کنید M بردار T_1, T_2, \dots, T_M ویژگی مکان مشخصه مجموعه داده های آموزشی هستند. هر ویژگی توسط برداری به طول $n=8200$ ارائه می شود. بردار میانگین، A ، و اختلاف بردارهای ویژگی از بردار میانگین،

X_i ، از روابط (1) و (2) محاسبه می شوند.

$$A = \frac{1}{M} \sum_{n=1}^M T_n \quad (1)$$

$$X_i = T_i - A \quad 1 < i < M \quad (2)$$

اگر ماتریس Y را توسط رابطه (3) تعریف کنیم، ماتریس کوواریانس C به کمک رابطه (4) محاسبه خواهد شد.

$$Y = [X_1 \quad X_2 \quad \dots \quad X_M]^T \quad (3)$$

$$C = \frac{1}{M} \sum_{n=1}^M X_n X_n^T = \frac{1}{M} Y Y^T \quad (4)$$

ماتریس کوواریانس، یک ماتریس $N \times N$ است و تعیین N بردار ویژه و N مقدار ویژه به جهت سایز بزرگ داده، غیرعملی است. اگر $M < N$ باشد می توان به

¹⁴ Landsat Multi-Spectral Scanner

استفاده قرار می گیرد.

استفاده از ضرایب وزن‌ها به صورت دینامیک نیز کارایی خوبی در بعضی مسائل واقعی و مصنوعی نشان داده است [۴].

در این مقاله از روش ترکیب وزندار استفاده شده است. این ساختار رفتار بسیاری از سیستم‌های ترکیبی را در بر می گیرد. ایده اصلی این است که طبقه بند بعدی، خطاهای قبلی را بر طرف کند. در این ساختار، طبقه بند های پایه به عنوان level-0 و ترکیب کننده به عنوان level-1 قرار می گیرند. این ساختار در شکل ۱ آمده است.

در فاز آموزش، level-0 بایستی به صورت متنوع آموزش ببیند که می توان از تکنیک leave-one-out و یا تکنیک های دیگر، جهت کاهش همبستگی طبقه بندهای پایه استفاده نمود. یادگیری طبقه بند level-1 به گونه ای است که خطای هر یک از طبقه بند های level-0 برای نمونه های مختلف ورودی را اصلاح نماید. در فاز آموزش، الگو به طبقه بند های level-0 داده شده و بردار بدست آمده از خروجی این طبقه بند ها، به عنوان ورودی، به طبقه بند level-1 اعمال می شود. در خروجی این طبقه بند که به عنوان ترکیب کننده عمل می کند تصمیم گیری نهایی حاصل می شود.

۵- نتایج تجربی

برای بررسی کارایی روش پیشنهادی، در مسائل طبقه بندی نتایج سه طبقه بند ترکیب شد. از شبکه عصبی پرسپترون چند لایه به عنوان طبقه بند پایه استفاده شده است. سه طبقه بند با استفاده از الگوریتم یادگیری تندترین شیب، با وزنهای اولیه متفاوت و ساختارهای [۶:۱۹:۳۶] و [۶:۲۰:۳۶] و [۶:۲۱:۳۶] عمل طبقه بندی اولیه را انجام می دهند. نرخ طبقه بندی این طبقه بندهای پایه، به طور متوسط بعد از پنج دفعه آزمایش برای هر ساختار، به ترتیب ۸۶/۸۳٪ و ۸۶/۴۸٪ و ۸۶/۶۶٪ است.

ساختار کلی سیستم طبقه بند در شکل ۱ آمده است. چنانکه شکل ۱ نشان می دهد، این ساختار شامل سه مرحله است. در مرحله اول کاهش افزونگی بردار ویژگی استخراج شده از

همانطور که در بخش قبل توضیح داده شد، تحلیل مولفه های اصلی یکی از روشهای کلاسیک است که موجب کاهش ابعاد نمونه ها می شود بطوری که عمده اطلاعات اصلی حفظ شود [۵]. این مجموعه داده شامل عکس هایی با ابعاد ۸۲×۱۰۰ پیکسل می باشد که ۳۶ مولفه اول حاصل از PCA دارای همه مشخصات مهم مجموعه داده می باشد [۶]. این مجموعه داده دارای ابعاد ۴۲۹۰×۳۶ برای فاز آموزش و ۲۱۴۵×۳۶ برای فاز آزمایش به عنوان ورودی و نیز ابعاد ۴۲۹۰×۶ برای فاز آموزش و ۲۱۴۵×۶ برای فاز آزمایش به عنوان خروجی میباشد.

۴- روش ترکیب طبقه بند ها

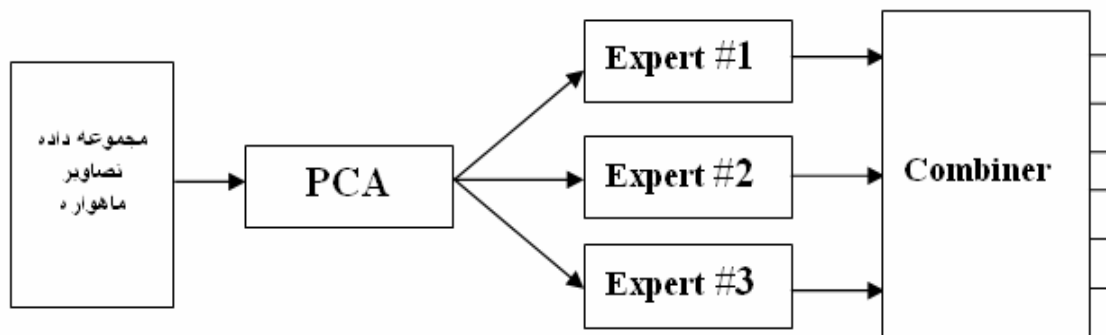
روش های متفاوتی برای ترکیب طبقه بند ها پیشنهاد شده است. از جمله این روش ها می توان رای گیری^{۱۵}، قوائد بیشینه، کمینه، میانگین، حاصلضرب، انتگرال فازی، ماتریس الگوی تصمیم و دمپستر-شفر را نام برد [۷ و ۸]. روش های فوق را می توان به صورت بدون وزن و وزندار استفاده نمود. حالت بدون وزن زمانی مورد استفاده قرار می گیرد که ارزش نظر تمام طبقه بند ها یکسان باشد و همه طبقه بند ها دارای کارایی یکسان بر روی همه کلاسها باشند. در این صورت $W_{ij} = W_j$ است. در حالت وزندار، کارایی هر طبقه بند در قسمت های مختلف فضای ویژگی، متفاوت است و لذا وزنهای تخصیص داده شده به کلاس های مختلف متفاوت خواهد بود. در این حالت نحوه تخصیص وزن ها به کلاسها و طبقه بند های مختلف اهمیت زیادی دارد. یک روش به منظور یافتن وزنهای بهینه، استفاده از روش حداقل میانگین مربعات خطاست [۹].

در صورتی که تعداد طبقه بند ها کم باشد، یک جستجوی دقیق می تواند برای یافتن وزنهای مفید باشد. روش دیگر استفاده از جستجوی ژنتیکی برای یافتن وزنهای می باشد.

¹⁵ Voting

نتایج طبقه بندی اولیه به شبکه عصبی لایه دوم (ترکیب کننده) اعمال می شود و توسط این شبکه عصبی نتایج لایه اول ترکیب شده و طبقه بندی نهایی صورت می گیرد.

داده های تصاویر ماهواره صورت می گیرد. در مرحله دوم، این بردارهای ویژگی به طبقه بند های پایه لایه اول اعمال شده و طبقه بندی اولیه صورت می گیرد. در مرحله سوم



شکل ۱ - سیستم پیشنهادی برای طبقه بندی دیتای تصویر ماهواره (Satimage)

۶- نتیجه گیری

با تلفیق نتایج طبقه بندهای مختلف می توان علاوه بر افزایش نرخ طبقه بندی، قدرت تعمیم سیستم تشخیص الگو را افزایش داد. نرخ طبقه بندی درست روی نمونه های تست برابر $88/413\%$ است. در حالی که بیشترین نرخ طبقه بندی تا کنون $86/1\%$ مربوط به مقاله Sharma و همکارانش [۱۰] بوده است که کارایی روش ارائه شده را نشان میدهد. البته باید توجه داشت که تلفیق نتایج خروجی چند طبقه بند، زمانی نتایج خوبی را حاصل می کند که طبقه بند ها دارای نرخ خطای کمی باشند و علاوه بر آن خطاهای آنها با یکدیگر متفاوت باشد یعنی الگوهایی که توسط یک طبقه بند بدرستی طبقه بندی نشده اند، توسط طبقه بند دیگری به صورت صحیح طبقه بندی شوند. این نکته در ساختار پیشنهادی با استفاده از شبکه های عصبی با وزنه های اولیه و تعداد نرونهای لایه میانی متفاوت محقق شده است.

از مجموعه داده های ELENA، مجموعه داده تصاویر ماهواره، شامل ۶۴۳۵ نمونه میباشد که ۴۲۹۰ نمونه برای آموزش شبکه و ۲۱۴۵ نمونه برای آزمایش سیستم استفاده شده است. بردار ویژگی استخراج شده دارای ۸۲۰۰ مولفه است. در مرحله کاهش افزونگی بردار ویژه به یک بردار ۳۶ مولفه ای تبدیل می شود. این بردار به عنوان ورودی شبکه عصبی (طبقه بند پایه) استفاده شد. طبقه بندهای پایه با استفاده از بردار ویژه گوی ۳۶ مولفه ای بر روی ۴۲۹۰ نمونه آموزشی تعلیم دیدند. پس از اتمام فاز آموزش طبقه بندهای پایه، نمونه های آموزشی مجدداً به طبقه بندهای پایه اعمال شد. در این مرحله خروجی طبقه بند های پایه یک بردار ۱۸ مولفه ای را تشکیل می دهند. این بردار به عنوان بردار ورودی ترکیب کننده محسوب شده و ترکیب کننده بر روی نمونه های آموزشی یادگیری حاصل می کند. در مرحله آزمایش بردار ویژگی مربوط به ۲۱۴۵ نمونه به سیستم اعمال شده و بردار خروجی ترکیب کننده مقادیر تعلق الگوی ورودی به کلاسهای مختلف را می دهد. کلاس با بیشترین مقدار تعلق به عنوان کلاس الگوی ورودی در نظر گرفته می شود.



Artificial Intelligence Research 11, 169–198, 1999.

مراجع :

[9]-س.ح. نبوی کریزی، ر. ابراهیم پور، ا. کبیر،
"کاربرد ترکیب طبقه بندها در بازشناسی ارقام
دستنویس فارسی"، سومین کنفرانس ماشین بینایی و
پردازش تصویر ایران، ص. 115 تا 119، اسفند 83.

[10]-A.Sharma, K. Paliwal, Godfrey C. Onwubolu "Class-dependent PCA,MDCand LDA."A combined classifier for patterclassification", pattern Recognition 39 (2006) 1215 – 1229

[1]-G. Sigletos , G. Paliouras, C.D. Spyropoulos."Combining Information Extraction Systems Using Voting and Stacked Generalization",Journal of Machine Learning Research 6 (2005) 1751-1782.

[2] - C.Nadal, R. Legault and C.Y. Suen. "Complementary Algorithms for Recognition of totallyUnconstrained Handwritten Numerals", Proc. 10th Int. Conf. Pattern Recognition, Vol. A, pp. 434-449, 1990.

[3] - J. Kittler, S. A. Hojjatoleslami, T. Windeatt." Weighting Factors in Multiple Expert Fusion", In Proc. Of British Machine vision Conference BMVC 97, pp. 42-50, Essex University, Essex U.K.,1997.

[4] - T.Windeatt , R.Ghaderi , " Dynamic Weighting Factors for Decision Combining",Proc. of IEE Int. Conf. On Data Fusion, Great Malvern, UK, pp. 123-130, October 1998.

[5]-R.O.Duda, and Hart, P.E.,Pattern Classification and Scene Analysis,John Wiley & Sons, 1973.

[6]- N. Wanas, Feature Based Architecture for Decision Fusion, thesis requirement for the degree of Doctor of Philosophy, Waterloo, Ontario, 2003.

[7] - E.Bauer, R.Kohavi , " An empirical comparison of voting classification algorithms: Bagging,boosting, and variants, " Mach. Learning J. 36, 105–139, 1999.

[8] - D.Opitz,, R.Maclin, , "Popular ensemble methods: A survey",Journal of