

مروری کوتاه بر یادگیری جمعی

علیرضا شبانی

گروه مهندسی صنایع، دانشکده‌ی مهندسی، دانشکدگان فارابی، دانشگاه تهران، تهران، ایران

چکیده

در دنیای امروز از یادگیری ماشین برای حل مسائل گوناگون بطور قابل توجهی استفاده می‌شود. یادگیری جمعی یک حوزه‌ی بسیار مهم در یادگیری ماشین می‌باشد. در سال‌های اخیر یادگیری جمعی توجهات زیادی را در زمینه‌ی هوش مصنوعی و داده‌کاوی به خود جلب کرده است. یادگیری جمعی نشان داده که در سطح گسترده‌ای از مسائل دنیای واقعی بسیار کارآمد و کاربردی خواهد بود. در این روش چندین یادگیر پایه با یکدیگر ترکیب شده و خروجی هر کدام از طبقه‌بندها باهم ادغام می‌شوند تا یک مدل قوی ساخته شود. در مقایسه با روش‌های یادگیری ماشین واحد، دقت روش‌های یادگیری جمعی بهبود چشم‌گیری دارد. هدف این مقاله، بررسی تکنیک‌های رایج یادگیری جمعی و ارائه‌ی یک نمای کلی از این تکنیک‌ها است. ارزیابی نقاط قوت و ضعف آن‌ها از منظر کاربردی و بینش‌های بدست آمده از این بررسی‌ها به محققان در انتخاب مناسب الگوریتم‌های یادگیری جمعی بطوریکه با نیازهایشان هم‌راستا باشد، کمک می‌کند.

واژگان کلیدی: یادگیری ماشین، یادگیری جمعی، بگینگ، بوستینگ، استکینگ

مقدمه

روش‌های یادگیری جمعی بر اساس اصل تصمیم‌گیری جمعی عمل می‌کنند که در آن گروهی از کلاس‌بندهای فردی برای تعیین مناسب‌ترین خروجی با یکدیگر ترکیب می‌شوند. تصمیم نهایی می‌تواند از طریق رای‌گیری یا میانگین احتمالی بدست بیاید. در رویکرد رای‌گیری هر کلاس‌بند پایه یک پیش‌بینی درمورد کلاس یک نمونه انجام می‌دهد و پیش‌بینی نهایی بر اساس تعداد رای‌های کلاس‌بندهای پایه به هر کلاس برای نمونه موردنظر انجام خواهد شد. در مورد رای‌گیری توصیه این است که تعداد کلاس‌بندها فرد درنظر گرفته شود تا حالت تساوی رخ ندهد. در رویکرد احتمالی هر کلاس‌بند پایه یک پیش‌بینی احتمالی درمورد کلاس نمونه موردنظر انجام می‌دهد و در نهایت میانگین پیش‌بینی مجموع کلاس‌بندها نتیجه‌ی نهایی را مشخص خواهد کرد. بدین ترتیب در اصطلاح به رویکرد اول رای‌گیری سخت و به رویکرد دوم رای‌گیری نرم گفته می‌شود [1-6].

ما از یادگیری جمعی در زندگی روزمره خود استفاده می‌کنیم. به عنوان یک مثال ساده، قبل از تماشای یک فیلم، برای اطمینان از ارزش فیلم، انتقادات و رتبه‌بندی فیلم که بر اساس تصمیم جمعی افرادیست که قبلاً فیلم را تماشا کرده‌اند بدست آمده را بررسی می‌کنیم. یک روش یادگیری جمعی نیز از همین مفهوم برای پیش‌بینی صحیح استفاده می‌کند.

با توجه به افزایش تنوع در مدل‌های پایه، یادگیری جمعی مسئله‌ی بیش‌برازش را کاهش محسوسی می‌دهد. یادگیری جمعی با موفقیت در زمینه‌های مختلف بکار گرفته شده و از مدل‌های یادگیری ماشین واحد بهتر عمل کرده است [1-6].

الگوریتم‌های یادگیری جمعی کاربردهایی را در زمینه‌های مختلف مراقبت‌های پزشکی و سلامت [8، 7]، تشخیص چهره [9]، تشخیص تقلب‌های مالی [10]، تحلیل احساسات [11] و بطور کلی تمامی زمینه‌هایی که یادگیری ماشین در آن کاربرد دارد، ارائه می‌دهند. سه تکنیک اصلی یادگیری جمعی که از نظر نحوه آموزش و ترکیب مدل‌های پایه متفاوت هستند وجود دارد که عبارتند از بگینگ، بوستینگ و استکینگ [1-6]. در بخش‌های بعدی مروری بر این سه تکنیک که بطور گسترده استفاده می‌شوند، خواهد شد.

محدودیت‌های روش‌های یادگیری ماشین واحد

برای ایجاد یک جمع دقیق از کلاس‌بندها، اطمینان از اینکه مجموع کلاس‌بندها عملکرد بهتری نسبت به هر کلاس‌بند واحد تشکیل‌دهنده جمع دارد، ضروری است. دقت هر کلاس‌بند استفاده شده در جمع در مقایسه با پیش‌بینی تصادفی (که دارای احتمال ۰.۵ می‌باشد) باید بهتر باشد. سه دلیل اصلی که موجب توسعه‌ی روش‌های یادگیری جمعی در مقابل الگوریتم‌های یادگیری ماشین واحد شد شامل موارد زیر می‌باشند [6، 12]:

۱. **مسئله آماری:** هدف الگوریتم‌های یادگیری ماشین معمولاً شناسایی بهترین فرضیه‌ها در یک فضای فرضیه معین است. با این حال، زمانی که مقدار ناکافی داده آموزشی وجود داشته باشد یا مجموعه داده آموزشی در مقایسه با اندازه فضای فرضیه نسبتاً کوچک باشد، یک مشکل آماری ظاهر می‌شود. در این سناریو، الگوریتم یادگیری ماشین ممکن است با فرضیه‌های متعددی در فضای فرضیه مواجه شود که سطح دقت یکسانی را به همراه دارد. الگوریتم‌های یادگیری جمعی در چنین شرایطی به کمک می‌آیند. آن‌ها این چالش آماری را با تجمیع پیش‌بینی‌ها یا آرا از طبقه‌بندی‌کننده‌های متعدد برطرف می‌کنند. این فرآیند تجمیع، احتمال انتخاب یک طبقه‌بندی نادرست را کاهش می‌دهد و متعاقباً دقت پیش‌بینی‌های داده‌های آموزشی را افزایش خواهد داد. در اصل، الگوریتم‌های جمعی به کاهش مشکل آماری ناشی از مقدار محدود داده‌های آموزشی با ترکیب خروجی‌های طبقه‌بندی‌کننده‌های متعدد برای پیش‌بینی‌های قابل اعتمادتر کمک می‌کنند.

۲. **مسئله محاسباتی:** الگوریتم‌های یادگیری ماشین با انجام نوعی جستجوی محلی برای شناسایی بهترین راه‌حل یا فرضیه عمل می‌کنند. با این حال، در موارد خاص، این الگوریتم‌ها ممکن است در بهینه‌ی محلی به دام بیفتند، حتی زمانی که مقدار زیادی از داده‌های آموزشی در دسترس باشد. قابل ذکر است، برای الگوریتم‌های درخت تصمیم و شبکه‌های عصبی، یافتن پیکربندی آموزشی

بهینه، یک مسئله NP-Hard است [13]. این بدان معنی است که تعیین بهترین فرضیه برای این مدل‌ها از نظر محاسباتی چالش برانگیز است. یادگیری جمعی می‌تواند این مشکل محاسباتی را کاهش دهد. با ایجاد مجموعه‌ای از یادگیران و اجرای جستجوهای محلی از نقاط شروع مختلف، رویکرد جمعی می‌تواند احتمال تقریب عملکرد دقیق و در عین حال ناشناخته را به طور موثرتری در مقایسه با یک یادگیرنده پایه افزایش دهد.

۳. مسئله بازنمایی: در بسیاری از موارد یادگیری ماشین، ممکن است نتوان فضای فرضیه را با استفاده از یک تابع واقعی به طور دقیق نشان داد. برای رفع این محدودیت، الگوریتم‌های یادگیری ماشین از مفهوم ترکیب وزنی از فرضیه‌های مختلف استفاده می‌کنند و در نتیجه دامنه توابع قابل بیان را گسترش می‌دهند. این بدان معناست که با اختصاص وزن‌های مختلف به فرضیه‌های مختلف، نمایش مجموعه وسیع‌تری از توابع ممکن می‌شود. الگوریتم‌هایی مانند شبکه‌های عصبی و درخت تصمیم می‌توانند کل فضای جستجوی طبقه‌بندی‌کننده‌ها را زمانی که داده‌های آموزشی فراوان هستند، کاوش کنند. بنابراین، داشتن فضای کارآمد و جامعی از فرضیه‌ها ضروری است که الگوریتم یادگیری بتواند آن را برای نمایش مناسب روابط پیچیده در داده‌ها بررسی کند.

یک الگوریتم یادگیری ماشین که در مسئله آماری ضعف دارد، دارای واریانس بالا می‌باشد. در صورتیکه در مسئله محاسباتی دچار ضعف باشد، دارای واریانس محاسباتی بالا خواهد بود. همچنین اگر در مسئله بازنمایی ضعف خود را نشان دهد، احتمالاً دارای بایاس بالا است. بنابراین از طریق الگوریتم‌های یادگیری جمعی، واریانس و بایاس الگوریتم‌های یادگیری واحد متعادل خواهند شد [6, 12].

روش‌های یادگیری جمعی

برای بهبود نتایج یادگیری ماشین واحد می‌توان از الگوریتم‌های یادگیری جمعی استفاده کرد. سه الگوریتم اصلی و مطرح یادگیری جمعی عبارتند از: بگینگ، بوستینگ و استکینگ. هدف اصلی پژوهش حاضر، بررسی اصول اولیه و ویژگی‌های مرتبط با هر یک از روش‌های جمعی می‌باشد. در ادامه درمورد نحوه کار هر روش، نحوه تولید داده‌ها، نحوه آموزش کلاس‌بندهای پایه و روش‌های ترکیب نتایج هر طبقه‌بند در هر کدام از روش‌ها بحث خواهد شد.

۱. بگینگ (Bagging) [14]: اصطلاح Bagging از ترکیب Bootstrap و Aggregating به وجود آمده است. این روش توسط بریمن در سال ۱۹۹۶ معرفی شد. Bagging نشان دهنده‌ی یک رویکرد الگوریتمی داده محور است. فرض اصلی آن حول ایجاد چندین زیر مجموعه مجزا از مجموعه داده اصلی می‌چرخد. هدف Bagging افزایش تنوع مدل‌های پیش‌بینی با ایجاد مجموعه داده‌های آموزشی تصادفی است، که تغییرات جزئی در داده‌های آموزشی می‌تواند منجر به تغییرات قابل توجهی در پیش‌بینی‌های مدل شود. از طرف دیگر، نتیجه‌ی نهایی از ترکیب نتایج کلاس‌بندهای پایه با روش رای‌گیری اکثریت مشخص خواهد شد. هرچند که می‌توان از انواع روش‌های رای‌گیری استفاده نمود.

این الگوریتم با کاهش واریانس به طور موثری مشکل بیش‌برازش را نیز کاهش می‌دهد که به ویژه در مجموعه داده‌های با ابعاد بالا کاربرد فراوان خواهد داشت. با این وجود البته این روش هزینه‌ی محاسباتی بالایی دارد و همچنین در مواردی دچار خطای سوگیری می‌باشد.

روش Bagging در تعیین پارامترهای بهینه مانند تعداد کلاس‌بندهای پایه چالش برانگیز است. همچنین برای رسیدن به نتایج با دقت بالاتر گاهی لازم است انواع روش‌های ترکیب خروجی‌های کلاس‌بندهای پایه را پیاده‌سازی و بررسی کرد.

الگوریتم جنگل تصادفی [15] یک توسعه از Bagging می‌باشد. تفاوت اولیه آن از Bagging در انتخاب ویژگی تصادفی در فرایند آموزش مدل است. در این روش با انتخاب تصادفی دسته‌ای از ویژگی‌ها، تنوع درخت‌های تصمیم افزایش می‌یابد که در مقابل Bagging که

همه‌ی ویژگی‌ها را در فرایند آموزش مدل دخیل می‌کند، دقت بالاتری ارائه می‌دهد.

روش درخت تصادفی متغیر (VR-Tree) [16] نیز روشی برای تولید درخت‌های تصمیم تصادفی است. این رویکرد شامل یک فرآیند انتخاب ویژگی و تعیین روش تقسیم درخت می‌باشد.

الگوریتم جنگل ایزوله (iForest) [17] نیز بویژه در تشخیص داده‌های پرت کاربرد دارد. این الگوریتم یک ویژگی (بُعد) را بصورت تصادفی انتخاب کرده و سپس یک مقدار تصادفی بین کمینه و بیشینه‌ی آن انتخاب می‌کند و در نهایت با یک خط جداساز آن بُعد را جدا خواهد کرد.

۲. بوستینگ (Boosting): این روش توسط فروند و شاپر در سال ۱۹۹۷ معرفی شد [18]. Boosting یک فرآیند یادگیری متوالی است که با تصحیح خطاهای ایجاد شده توسط مدل‌های قبلی به یک پیش‌بینی با دقت بالا می‌رسد. این روش از چند یادگیر پایه تشکیل شده که با تمرکز بر خطاهای مدل پیشین به تدریج عملکرد خود را بهبود می‌بخشد. روش‌های ترکیب نتایج نیز متنوع است. Boosting هم برای مسائل کلاس‌بندی و هم مسائل رگرسیون کاربرد دارد.

این الگوریتم واریانس و بایاس را کاهش می‌دهد که در نتیجه افزایش دقت را به همراه خواهد داشت. Boosting نسبت به Bagging مدت‌زمان طولانی‌تری برای آموزش مدل صرف می‌کند و با بالا رفتن تعداد تکرارها در برابر بیش‌برازش آسیب‌پذیر می‌شود. همچنین این روش هزینه محاسباتی بالایی از خود نشان داده است.

الگوریتم AdaBoost [19] یکی از اصلی‌ترین روش‌های مربوط به Boosting می‌باشد که تاکید خاص آن بر مدیریت وزن‌های اختصاص داده شده به هر نمونه‌ی آموزشی است. در هر تکرار، الگوریتم وزن نمونه‌هایی را که به اشتباه طبقه‌بندی شده‌اند، افزایش می‌دهد و در نتیجه توجه یادگیرنده پایه را به سمت چالش‌برانگیزترین نمونه‌ها در مجموعه داده آموزشی سوق خواهد داد. نتیجه نهایی از طریق اکثریت آرا طبقه‌بندی‌کننده‌های پایه به دست می‌آید.

در AdaBoost.M1 [20] که توسعه‌ای از AdaBoost است، از یادگیرنده‌های چندکلاس به جای کلاس‌بندهای دوکلاس استفاده می‌شود و کاربرد این روش نیز در مجموعه داده‌های چندکلاس می‌باشد.

AdaBoost.M2 [20] نیز برای مجموعه داده‌های چندکلاس بکار می‌رود. در این روش یک مسئله کلاس‌بندی چندکلاس به چند مسئله کلاس‌بندی دوکلاس تبدیل می‌شود. در نهایت مدل باید تشخیص دهد که یک نمونه به کلاس i تعلق دارد یا خیر.

۳. استکینگ (Stacking): این روش که در سال ۱۹۹۲ توسط اسمیت و ولپرت معرفی شد، با ترکیب اطلاعات بدست آمده از چند مدل و ایجاد یک متا-مدل به دقت بالایی دست پیدا می‌کند [21]. معماری یک مدل Stacking شامل دو یا چند مدل پایه است که به عنوان مدل سطح ۰ نامیده می‌شود و یک متا-مدل به عنوان مدل سطح ۱ که پیش‌بینی‌های تولید شده توسط مدل‌های پایه را باهم ترکیب می‌کند. در مدل‌های سطح ۰، مدل‌های فردی بر روی داده‌های آموزشی موجود آموزش داده می‌شوند و پیش‌بینی‌های مربوط به آن‌ها ذخیره می‌شود. سپس مدل سطح ۱ یا متا-مدل یاد می‌گیرد که چگونه پیش‌بینی‌های تولید شده توسط مدل‌های پایه را بطور بهینه ترکیب کند. پیش‌بینی‌های ارائه شده به متا-مدل می‌تواند بصورت مقادیر احتمالی یا قطعی باشد.

Stacking کارآمدی بسیاری در زمینه‌های مختلف از خود نشان داده است چرا که از هر مدل فردی به طرز قابل توجهی بهتر عمل می‌کند. از مزیت‌های کلیدی این روش، توانایی آن در ایجاد درک عمیق از داده‌ها است که در نتیجه دقت پیش‌بینی را به شدت افزایش می‌دهد.

این الگوریتم نیز بدون چالش نخواهد بود. از آنجایی که مدل‌های متعددی برای پیش‌بینی یک هدف مشترک تلاش می‌کنند، ممکن است که بیش‌برازش اتفاق بیفتد. علاوه بر این، این الگوریتم به داده‌های بیشتری برای آموزش نیاز دارد و همچنین آموزش آن زمان‌بر خواهد بود.

در پیاده‌سازی Stacking با مسائلی همچون تعیین تعداد بهینه‌ی مدل‌های پایه و انتخاب قابل اعتمادترین مدل‌های پایه برای بهینه‌سازی

پیش‌بینی‌ها مواجه خواهیم شد. در مسائل چندکلاسه، پیچیدگی این مدل افزایش می‌یابد. در پژوهشی ولپرت و همکاران بر اهمیت در نظر گرفتن ویژگی‌های مختلف برای مجموعه داده‌های مورد استفاده و دسته‌بندی الگوریتم‌های یادگیری مناسب برای متا-مدل تاکید کردند. الگوریتم SCANN یک توسعه از Stacking می‌باشد که از تجزیه و تحلیل همبستگی برای شناسایی روابط بین پیش‌بینی‌های انجام شده توسط طبقه‌بندی کننده‌های سطح پایه استفاده می‌کند. سپس این پیش‌بینی‌ها برای حذف وابستگی‌ها بازسازی می‌شوند و فضای ویژگی جدید حاصل، از نزدیک‌ترین همسایه به عنوان طبقه‌بندی کننده متا استفاده می‌کند [22]. روش Grading [23] نیز شامل ایجاد یک کلاس‌بند در سطح متا برای هر کلاس‌بند سطح پایه است که امکان ترکیب‌های پیچیده‌تر و متناسب‌تر از خروجی‌های کلاس‌بند را فراهم می‌کند.

بحث و نتیجه‌گیری

در این مقاله مفهوم مهمی به نام یادگیری جمعی مورد بحث قرار گرفت که از مباحث رایج در یادگیری ماشین است. سه الگوریتم اصلی یادگیری جمعی شامل Bagging، Boosting و Stacking بررسی شد و مزایا و معایب این الگوریتم‌ها به همراه کیفیت عملکرد آن‌ها نیز مورد مطالعه واقع شد. در کنار آن، نمونه‌هایی از توسعه‌های اصلی این الگوریتم‌ها مورد اشاره کلی قرار گرفت. انتظار می‌رود که این پژوهش درک و بینشی درست به خوانندگان بدهد تا تصمیمی آگاهانه در شناسایی گزینه‌های موجود در الگوریتم‌های یادگیری ماشین بگیرند و سپس الگوریتم یادگیری ماشین مناسب را در زمینه حل مسئله خاص انتخاب کنند. درک محاسن و معایب این الگوریتم‌ها، تصمیم‌گیری آگاهانه، تضمین عملکرد بهینه و اجرای موفقیت‌آمیز در سناریوهای دنیای واقعی را تسهیل خواهد کرد.

منابع

- [1] Polikar, R. (2006). Ensemble based systems in decision making. IEEE Circuits and systems magazine, 6(3), 21-45.
- [2] Polikar, R. (2012). Ensemble learning. Ensemble machine learning: Methods and applications, 1-34.
- [3] Rokach, L. (2010). Ensemble-based classifiers. Artificial intelligence review, 33, 1-39.
- [4] Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1249.
- [5] Rokach, L. (2019). Ensemble learning: pattern classification using ensemble methods.
- [6] Zhou, Z. H. (2019). Ensemble methods: foundations and algorithms. Chapman and Hall/CRC.
- [7] Jothi Prakash, V., & Karthikeyan, N. K. (2021). Enhanced evolutionary feature selection and ensemble method for cardiovascular disease prediction. Interdisciplinary Sciences: Computational Life Sciences, 13(3), 389-412.
- [8] AlJame, M., Ahmad, I., Imtiaz, A., & Mohammed, A. (2020). Ensemble learning model for diagnosing COVID-19 from routine blood tests. Informatics in Medicine Unlocked, 21, 100449.
- [9] Han, S., Meng, Z., Khan, A. S., & Tong, Y. (2016). Incremental boosting convolutional neural network for facial action unit recognition. Advances in neural information processing systems, 29.
- [10] Zareapoor, M., & Shamsolmoali, P. (2015). Application of credit card fraud detection: Based on bagging ensemble classifier. Procedia computer science, 48(2015), 679-685.
- [11] Fouad, M. M., Gharib, T. F., & Mashat, A. S. (2018). Efficient twitter sentiment analysis system with feature selection and classifier ensemble. In The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2018) (pp. 516-527). Springer International Publishing.
- [12] Dietterich, T. G. (2000, June). Ensemble methods in machine learning. In International workshop on multiple classifier systems (pp. 1-15). Berlin, Heidelberg: Springer Berlin Heidelberg.
- [13] Blum, A., & Rivest, R. (1988). Training a 3-node neural network is NP-complete. Advances in neural information processing systems, 1.
- [14] Breiman, L. (1996). Bagging predictors. Machine learning, 24, 123-140.
- [15] Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.
- [16] Liu, F. T., Ting, K. M., Yu, Y., & Zhou, Z. H. (2008). Spectrum of variable-random trees. Journal of Artificial Intelligence Research, 32, 355-384.
- [17] Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008, December). Isolation forest. In 2008 eighth IEEE international conference on data mining (pp. 413-422). IEEE.
- [18] Schapire, R. E. (1990). The strength of weak learnability. Machine learning, 5, 197-227.
- [19] Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences, 55(1), 119-139.
- [20] Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting". Journal of Computer and System Sciences, pp: 119–139, (2000).

- [21] Wolpert, D. H. (1992). Stacked generalization. *Neural networks*, 5(2), 241-259.
- [22] Merz, C. J. (1999). Using correspondence analysis to combine classifiers. *Machine Learning*, 36, 33-58.
- [23] Seewald, A. K., & Fürnkranz, J. (2001, September). An evaluation of grading classifiers. In *International symposium on intelligent data analysis* (pp. 115-124). Berlin, Heidelberg: Springer Berlin Heidelberg.

A brief review of ensemble learning

Alireza Shabani

Department of Industrial Engineering, Faculty of Engineering, College of Farabi, Universty of Tehran, Tehran, Iran

Abstract

In the contemporary landscape, machine learning plays a pivotal role in addressing a myriad of challenges. Ensemble learning emerges as a crucial domain within the realm of machine learning, garnering substantial interest in the spheres of artificial intelligence and data mining in recent years. Demonstrating notable efficiency and applicability across a diverse array of real-world scenarios, ensemble learning involves the amalgamation of multiple basic learners, with the outputs of individual classifiers combined to construct a robust model. Notably, ensemble learning methods exhibit a marked enhancement in accuracy compared to singular machine learning approaches. This article aims to examine prevalent ensemble learning techniques, offering a comprehensive overview of these methodologies. Through a pragmatic assessment of their strengths and limitations and the insights gleaned from such evaluations, researchers can make informed decisions regarding the selection of suitable ensemble learning algorithms that align with their specific requirements.

Keywords: Machine learning, Ensemble learning, Bagging, Boosting, Stacking