دانشگاه تربیت مدرس
دانشکده علوم پزشکی

پایان‌نامه
دوره کارشناسی ارشد رشته آمار زیستی

عنوان

شبکه عصبی مصنوعی با مدل خبره-آمیخته و کاربرد آن در تحلیل داده های پزشکی

دانشجو
مرضیه ابراهیمی

استاد راهنما:
دکتر انوشیروان کاظم نژاد

استاد مشاور:
دکتر محمد غلامی فشارکی

زمستان ۱۳۹۳

تاییدیه اعضای هیات داوران حاضر در جلسه دفاع از
پایان‌نامه کارشناسی ارشد

خانم مرضیه ابراهیمی رشته **آمار زیستی** پایان‌نامه **کارشناسی ارشد** خود را با عنوان « **شبکه عصبی مصنوعی با مدل خبره– آمیخته و کاربرد آن در تحلیل داده های پزشکی** »در تاریخ **۱۳۹۳/۱۲/۲۰** ارائه کردند.

بدینوسیله اعضای هیات داوران نسخه نهایی این پایان‌نامه را از نظر فرم و محتوا تایید کرده و پذیرش آنرا برای تکمیل درجه کارشناسی ارشد پیشنهاد می‌کنند.

اعضای هیات داوران	نام و نام خانوادگی	امضاء
استاد راهنما	دکتر انوشیروان کاظم نژاد	
استاد مشاور	دکتر محمد غلامی فشارکی	
استاد ناظر	دکتر فرید زایری	
استاد ناظر و نماینده تحصیلات تکمیلی	دکتر علی اکبر راسخی	

آیین‌نامه حق مالکیت مادی و معنوی در مورد نتایج پژوهشهای علمی
دانشگاه تربیت مدرس

مقدمه: با عنایت به سیاست‌های پژوهشی و فناوری دانشگاه در راستای تحقق عدالت و کرامت انسانها که لازمه شکوفایی علمی و فنی است و رعایت حقوق مادی و معنوی دانشگاه و پژوهشگران، لازم است اعضای هیأت علمی، دانشجویان، دانش‌آموختگان و دیگر همکاران طرح، در مورد نتایج پژوهشهای علمی که تحت عناوین پایان‌نامه، رساله و طرحهای تحقیقاتی با هماهنگی دانشگاه انجام شده است، موارد زیر را رعایت نمایند:

ماده ۱- حق نشر و تکثیر پایان نامه/ رساله و درآمدهای حاصل از آنها متعلق به دانشگاه می باشد ولی حقوق معنوی پدید آورندگان محفوظ خواهد بود.

ماده ۲- انتشار مقاله یا مقالات مستخرج از پایان‌نامه/ رساله به صورت چاپ در نشریات علمی و یا ارائه در مجامع علمی باید به نام دانشگاه بوده و با تایید استاد راهنمای اصلی، یکی از اساتید راهنما، مشاور و یا دانشجوی مسئول مکاتبات مقاله باشد. ولی مسئولیت علمی مقاله مستخرج از پایان نامه و رساله به عهده اساتید راهنما و دانشجو می باشد.

تبصره: در مقالاتی که پس از دانش‌آموختگی بصورت ترکیبی از اطلاعات جدید و نتایج حاصل از پایان‌نامه/ رساله نیز منتشر می‌شود نیز باید نام دانشگاه درج شود.

ماده ۳- انتشار کتاب و یا نرم افزار و یا آثار ویژه(اثری هنری مانند فیلم، عکس، نقاشی و نمایشنامه) حاصل از نتایج پایان‌نامه/ رساله و تمامی طرحهای تحقیقاتی کلیه واحدهای دانشگاه اعم از دانشکده ها، مراکز تحقیقاتی، پژوهشکده ها، پارک علم و فناوری و دیگر واحدها باید با مجوز کتبی صادره از معاونت پژوهشی دانشگاه و براساس آئین‌نامه‌های مصوب انجام شود.

ماده ۴- ثبت اختراع و تدوین دانش فنی و یا ارائه یافته ها در جشنواره‌های ملی، منطقه‌ای و بین‌المللی که حاصل نتایج مستخرج از پایان‌نامه/ رساله و تمامی طرح‌های تحقیقاتی دانشگاه باید با هماهنگی استاد راهنما یا مجری طرح از طریق معاونت پژوهشی دانشگاه انجام گیرد.

ماده ۵- این آیین‌نامه در ۵ ماده و یک تبصره در تاریخ۸۷/۴/۱ در شورای پژوهشی و در تاریخ ۸۷/۴/۲۳ در هیأت رئیسه دانشگاه به تایید رسید و در جلسه مورخ ۸۷/۷/۱۵ شورای دانشگاه به تصویب رسیده و از تاریخ تصویب در شورای دانشگاه لازم‌الاجرا است.

«اینجانب **مرضیه ابراهیمی** دانشجوی رشته **آمار زیستی** ورودی سال تحصیلی**۱۳۹۱**مقطع **کارشناسی ارشد** دانشکده علوم **پزشکی** متعهد می شوم کلیه نکات مندرج در آیین نامه حق مالکیت مادی و معنوی در مورد نتایج پژوهش های علمی دانشگاه تربیت مدرس را در انتشار یافته های علمی مستخرج از پایان نامه / رساله تحصیلی خود رعایت نمایم. در صورت تخلف از مفاد آیین نامه فوق الاشعار به دانشگاه وکالت و نمایندگی می دهم که از طرف اینجانب نسبت به لغو امتیاز اختراع بنام بنده و یا هرگونه امتیاز دیگر و تغییر آن به نام دانشگاه اقدام نماید. ضمناً نسبت به جبران فوری ضرر و زیان حاصله براساس برآورد دانشگاه اقدام خواهم نمود و بدینوسیله حق هرگونه اعتراض را از خود سلب نمودم ».

امضا

تاریخ

۹۴/۳/۵

آئین‌نامه پایان‌نامه (رساله)های دانشجویان دانشگاه تربیت مدرس

نظر به اینکه چاپ و انتشار پایان‌نامه (رساله)های تحصیلی دانشجویان دانشگاه تربیت مدرس، مبین بخشی از فعالیت‌های علمیپژوهشی دانشگاه است. بنابراین به منظور آگاهی و رعایت حقوق دانشگاه، دانش آموختگان این دانشگاه نسبت به رعایت موارد ذیل متعهد می‌شوند:

ماده ۱ : در صورت اقدام به چاپ پایان‌نامه (رساله)ی خود، مراتب را قبلاً به‌طور کتبی به دفتر "دفتر نشر آثار علمی" دانشگاه اطلاع دهد.

ماده ۲ : در صفحه سوم کتاب (پس از برگ شناسنامه)، عبارت ذیل را چاپ کند:

" کتاب حاضر، حاصل **پایان‌نامه کارشناسی ارشد** نگارنده در رشته **آمار زیستی** است که در سال **۱۳۹۳** در دانشکده علوم پزشکی دانشگاه تربیت مدرس به راهنمایی **دکتر انوشیروان کاظم نژاد** ، مشاوره **دکتر محمد غلامی** از آن دفاع شده است.

ماده ۳ : به منظور جبران بخشی از هزینه‌های انتشارات دانشگاه، تعداد یک درصد شمارگان کتاب (در هر نوبت چاپ) را به "دفتر نشر آثار علمی" دانشگاه اهداء کند. دانشگاه می‌تواند مازاد نیاز خود را به نفع مرکز نشر در معرض فروش قرار دهد.

ماده ۴ : در صورت عدم رعایت ماده ۳، ۵۰٪ بهای شمارگان چاپ شده را به‌عنوان خسارت به دانشگاه تربیت مدرس، تادیه کند.

ماده ۵ : دانشجو تعهد و قبول می‌کند در صورت خودداری از پرداخت‌های بهای خسارت، دانشگاه مذکور را از طریق مراجع قضایی مطالبه و وصول کند، به علاوه به دانشگاه حق می‌دهد به منظور استیفای حقوق خود، از طریق دادگاه، معادل وجه مذکور در ماده ۴ را از محل توقیف کتابهای عرضه شده نگارنده برای فروش، تامین نماید.

ماده ۶ : اینجانب **مرضیه ابراهیمی** دانشجوی رشته **آمار زیستی** مقطع **کارشناسی ارشد** تعهد فوق و ضمانت اجرایی آن را قبول کرده، به آن ملتزم می‌شوم.

نام‌و نام خانوادگی مرضیه ابراهیمی

تاریخ و امضا

۹۴/۳/۱۵

تقدیم

به پدر و مادر و خواهر عزیزم

آنان که باور ناب بودن، لذت زیستن، جسارت خواستن، عظمت رسیدن و تمام تجربه‌های یکتای زندگیم، مدیون حضور سبز شان است.

تشکر و قدردانی

شکر شایان نثار ایزد منان که توفیق رارفیق راهم نمود و به همنشینی رهروان علم و دانش مفتخرم ساخت. اکنون که پایان نامه حاضر به انجام رسیده است، بر خود لازم می دانم مراتب سپاس را از بزرگوارانی به جا آورم که اگر دست یاریگر شان نبود، هرگز این تحقیق با انجام نمی رسید.

از اولین و بهترین معلمان زندگیم پدر و مادر مهربانم که همیاری پر مهرشان دشواری راه را برایم هموار ساخت.

بسی شایسته است قدردانی کنم از استاد گرانقدر جناب آقای دکتر انوشیروان کاظم نژاد که زحمت راهنمایی این پایان نامه را بر عهده داشتند و همچنین از جناب آقای دکتر غلامی که همواره با راهنمایی های عالمانه و ارزشمند، مرا مورد لطف و محبت قرار داده اند.

از اساتید گرامی جناب آقای دکتر راسخی و جناب آقای دکتر زایری که زحمت داوری این پایان نامه را متقبل شدند، کمال سپاس را دارم.

با احترام

مرضیه ابراهیمی

چکیده

زمینه و اهداف: در تحقیقات پزشکی مدلسازی و پیش‌بینی از اهمیت زیادی برخوردار است. روش‌های پیش‌بینی را می‌توان با تکنیک هایی بهبود داد. مدل های آماری و شبکه عصبی مصنوعی از مدل‌هایی هستند که در رده‌بندی و پیش‌بینی مورد استفاده قرار می‌گیرد، اما مدل‌های آماری نیاز به پیش‌فرض-هایی دارند و شبکه عصبی نیازمند حجم نمونه کافی برای آموزش است.از این رو در این پایان‌نامه اختلاط خبره‌ها را معرفی می‌کنیم که یکی از مدل‌های رایج در ترکیب دسته‌بندها است. ترکیب دسته‌بندها روشی برای بهبود کارایی در مسائل رده‌بندی است که دارای تعداد الگوی محدودتری هستند. این مدل‌ها برای رگرسیون و رده‌بندی مفیدند، ولی ما تنها به دسته‌بندی توسط این مدل پرداخته‌ایم. در اختلاط خبره‌ها فضا بین چند دسته‌بند تقسیم می‌شود و با یک شبکه میانجی اجرا می‌گردد. اختلاط خبره‌ها با استفاده از روش‌های تکراری نظیر الگوریتم ماکسیمم مورد انتظار و شبکه های عصبی قابل برازش است.

روش بررسی: این مطالعه شامل ۲۱۳ بیمار هپاتیت سی (۱۹۵ مرد، ۱۸ زن، فاصله سنی ۱۲-۶۶) است. متغیر پاسخ در این مدل بهبودی پس از درمان در نظر گرفته شده‌است. در ابتدا مدل لجستیک به داده ها برازش داده شد و فاکتورهای موثر شناخته شده‌اند و بعد از آن مدل‌های شبکه عصبی و اختلاط خبره‌ها برازش داده شد. . برای مقایسه عملکرد این سه مدل از منحنی مشخصه عملکرد و صحت پیش‌بینی استفاده شده است.

یافته‌ها: نتایج حاصل از مدل لجستیک نشان‌دهنده معناداری متغیرهای شاخص توده بدنی، سن، بارویروسی، RS12 است. در تحلیل این داده‌ها با دو مدل لجستیک و شبکه عصبی،سطح پیشگویی مشابه‌ی را نشان داده است. مدل اختلاط خبره‌ها توسط شبکه عصبی برازش داده شد.این مدل در سطح عملکرد پیشگویی همانند شبکه عصبی بوده‌است. شبیه سازی‌ها این مطلب را تایید کردند که مدل اختلاط خبره‌ها یک مدل وابسته به ساختار داده است.

نتیجه: اختلاط خبره‌ها در مقایسه با یک مدل رده‌بندی عملکرد را بهبود می‌بخشد. اما این مدل به ساختار داده‌ها وابسته است.

کلید واژه: رده‌بندی، شبکه عصبی مصنوعی، اختلاط خبره‌ها، هپاتیت C.

فهرست مطالب

عنوان **صفحه**

فهرست شکل‌ها

عنوان **صفحه**

فهرست جداول

عنوان **صفحه**

فصل اول

مقدمه و مروری بر مطالعات گذشته

۱-۱ مقدمه

خلاصه‌ای از واقعیت را مدل گویند. به بیان دیگر، نمایش مجرّد یا فیزیکی یک سیستم یا سامانه را از یک دیدگاه و نگاه خاص مدلسازی می نامند. مدل‌ها، انواع گوناگونی مانند مدل فیزیکی، مدل ریاضی، مدل آماری، مدل گرافی، نرم‌افزاری، و... دارند و کاربردهای حیاتی متنوّع و فراوانی در همه زمینه‌های علوم و فناوری دارند. در این میان مدلسازی آماری نسبت به سایر انواع مدلسازی دارای تنوع بیشتری می باشد. روشهایی نظیر مدل‌های رگرسیونی(خطی و غیر خطی)، تحلیل خوشه‌ای، آنالیز تشخیصی، سری‌های زمانی و.... ازروشهای مدلسازی آماری هستند، در این این میان سه عامل نوع متغیرهای توضیحی، توزیع احتمالی متغیر پاسخ و رابطه موجود میان متغیرهای کمکی با متغیر پاسخ از عوامل مهم برانتخاب نوع مدل می باشد. مثلا وقتی با پاسخ های پیوسته مواجه هستیم رگرسیون خطی معمولی و یا مدل‌های تحلیل واریانس می‌توانند روش‌های مناسبی برای پیشبرد اهداف مطالعه باشند. حال هر چه متغیر پاسخ حساس تر باشد دقت پیش‌بینی صحیح نیز از اهمیت بیشتری بر خوردار می‌شود. بخصوص در حیطه علوم پزشکی و بهداشتی که در آن سلامت افراد مدنظر محققان می باشد. از این رو محققان همواره بدنبال روش‌های دقیقتر برای پیش‌بینی بوده‌اند که از آن جمله می توان به روش شبکه عصبی مصنوعی اشاره نمود.

شبکه‌های عصبی مصنوعی یا به زبان ساده‌تر شبکه‌های عصبی سیستم‌ها و روش‌های محاسباتی نوینی هستند برای یادگیری ماشینی، نمایش دانش، و در انتها اعمال دانش به دست آمده در جهت پیش‌بینی پاسخ‌های خروجی از سامانه‌های پیچیده است. ایده اصلی این گونه شبکه‌ها (تا حدودی) الهام‌گرفته از شیوه کارکرد سیستم عصبی زیستی، برای پردازش داده‌ها، و اطلاعات به منظور یادگیری و ایجاد دانش قرار دارد. عنصر کلیدی این ایده، ایجاد ساختارهایی جدید برای سامانه پردازش اطلاعات است. این سیستم از بی شمار عناصر پردازشی فوق‌العاده بهم‌پیوسته با نام نورون تشکیل شده که برای حل یک مسئله با هم هماهنگ عمل می‌کنند و توسط سیناپس‌ها (ارتباطات الکترومغناطیسی) اطلاعات را منتقل می‌کنند. در این شبکه‌ها اگر یک سلول آسیب ببیند بقیه سلول‌ها می‌توانند نبود آنرا جبران کرده، و نیز در بازسازی آن سهیم باشند. این شبکه‌ها قادر به یادگیری‌اند. یادگیری در این سیستم‌ها به صورت تطبیقی صورت می‌گیرد، یعنی با استفاده ازمثال‌ها وزن سیناپس‌ها به گونه‌ای تغییر می‌کند که در صورت دادن ورودی‌های جدید، سیستم پاسخ درستی تولید کند. توافق دقیقی بر تعریف شبکه عصبی در میان محققان وجود ندارد؛ اما اغلب آنها موافقند

که شبکه عصبی شامل شبکه‌ای از عناصر پردازش ساده (نورونها)است، که می‌تواند رفتار پیچیده کلی تعیین شده‌ای از ارتباط بین عناصر پردازش و پارامترهای عنصر را نمایش دهد. مطالعات نشان داده اند که شبکه عصبی برای تخمین[1] و تقریب[2] کارایی بسیار بالایی دارند. در برازش شبکه های عصبی مصنوعی از الگوریتم تکراری تا رسیدن به مینیمم خطا استفاده می‌شود. یکی از مشکلات استفاده از این الگوریتم گرفتار شدن در مینیمم های محلی تابع خطا است. بمنظور رفع این مشکل می‌توان از سیستم‌های شورایی استفاده کرد.

در جوامع برای تصمیم‌گیری‌های مهم در مواقع کلیدی مردم به گروهی که در مجلس شورا و یا هیاتی که مدیریت را بعهده گرفته اند اعتماد می‌کنند و این مجالس بر اساس این تفکر که دو فکر بهتر از یک فکر عمل می‌کند شکل می‌گیرند. در سال ۱۸۱۸ لاپلاس نشان می‌دهد که ترکیب مناسبی از دو روش احتمالاتی بهتر از مدلی با یک مؤلفه انجام می‌شود. روش هایی که از ترکیب چند دسته‌بند تشکیل می‌شوند به روش های شورایی معروف هستند[۱].

یکی از روش های جالب شورایی[3] ،اختلاط خبره ها[4] است که توان بالقوه در اجرای ماشین‌های یادگیری ایجاد می‌کند. این مدل بر اساس اصل تقسیم و تسخیر[5] عمل می‌کنند[۲]. در این مدل فضای مسئله بین چند خبره[6] با نظارت مدخل شبکه[7] تقسیم می‌شود. این مدل در سال ۱۹۹۱ توسط جیکوبز و جردن معرفی شده است[۲].

اختلاط خبرها ابتدا در رشته های کامپیوتر و الکترونیک به کار برده شد و سپس در سیستم های هوشمند[8] و سیستم‌های تصمیم‌گیری[9] در شبکه عصبی و تحلیل و پردازش تصاویر و اطلاعات کاربرد داشته است و از دیگر کاربرد های آن در علوم پزشکی به منظور تشخیص و پیش‌بینی میزان خطر بروز یک بیماری را می‌توان نام برد. از کاربردهای این روش در سال‌های اخیر در مطالعه در تشخیص بیماری دیابت و تشخیص سیروزکبدی و همچنین در مطالعاتی بر روی سرطان سینه قابل ذکر است[۴و۵]. هریس و همکارانش شبکه اختلاط خبره‌ها را برای رسیدن به کارایی بالاتر در مدل های رگرسیونی تعمیم داده اند. آنان معیار توقف را رسیدن به کمترین خطای کلی در شبکه قرار داده اند[۶].

---

[1] Estimation
[2] Approximation
[3] Ensamble
[4] Mixture of experts

منظور از خبره آمیخته همان اختلاط خبرها است.

[5] Divided and conqure
[6] Experts
[7] Gatting network
[8] Intelligent system
[9] Decision support system

۱-۲ اهداف پایان نامه

هدف از انجام این پایان‌نامه، به کارگیری مدل اختلاط خبره‌ها درپیش‌بینی داده ها است. این مدل، معمولاً جهت رده‌بندی داده ها با حجم کم بکار رفته و قابل ذکر است که صحت پیش‌بینی این مدل نسبت به یک مدل رده‌بندی افزایش چشمگیری را درمقالات نشان داده استو از این رو، اجرای آن بر روی داده‌هایمورد نظر انجام گرفته است. در این مطالعه به بررسی پارامترهای مختلف تاثیرگذار بر مدل پرداخته شده است.

۱-۳ نگاهی گذرا به پایان نامه

در این فصل یادگیری ماشین و ضروریات در انجام این روش بررسی شده است و همچنین داده های مورد استفاده در پایان نامه معرفی می‌شوند.

فصل دوم به بیان روش‌های رده‌بندی داده ها پرداخته شده است و نحوه تحلیل نتایج در مدل-های رده‌بندی و اعتبار سنجی دسته‌بندها[۱] معرفی می‌شوند، از روشهای متداول رده‌بندی می‌توان به آنالیز تشخیصی[۲] و مدل‌های لجستیک[۳] و شبکه‌های عصبی مصنوعی[۴] اشاره داشت که با توجه به نوع مسأله از توانمندی‌های متفاوتی برخوردارند. شبکه عصبی یکی از متداولترین ابزار های دسته بندی و پیش‌بینی در یادگیری ماشین است.این روش به عنوان مدل اصلی در مطالعه حاضر، به طور مفصل پرداخته شده است. برآورد پارامترها در شبکه عصبی با استفاده از الگوریتم یادگیری تکراری محاسبه می‌شود، این الگوریتم در جستجوی کمترین خطای شبکه هستند که ممکن است در کمینه های محلی گرفتار شود. یادگیری شورایی[۵] روشی است که به منظور بهبود دقت مدل رده‌بندی استفاده می‌گردد. بنابراین در ادامه به معرفی سیستم‌های دسته‌بند چندگانه[۶] پرداخته، و روش های برآورد این مدل ها بیان شده است.

در اختلاط خبره ها، هر مدل دسته‌بندی پایه با توجه به مقدار پارامترهایش، به پاسخ متفاوتی برای مسأله می‌رسد و با ترکیب پاسخ‌ها، دقت مدل افزایش میابد. به‌منظور ترکیب مناسبی از خروجی چند مدل، ابتدا باید در جستجوی چارچوب ریاضی برای ترکیب این مدل‌ها بود تا از نقاط قوت آنها استفاده شود و از نقاط ضعف آنها پرهیز گردد، همچنین طبق تحقیقات انجام گرفته ، وجود دسته‌بند های پایه با کارایی مناسب و تقریبا مستقل از یکدیگر از ملزومات این مدل می‌باشد [۷].

در متون بررسی شده از روش های آموزش مختلفی در مدل اختلاط خبره ها بهره گرفته شده

---

[1] Classifier
[2] Discriminant analysis
[3] Logestic
[4] Artificial neural network
[5] Ensamble learning
[5] Multiple classifier systems

است. از جمله روش‌های آموزش این سیستم‌ها، استفاده از الگوریتم ماکسیمم مورد انتظار[1] می باشد که در بعضی مقالات از این الگوریتم برای برآورد پارامترهای مدل اختلاط خبره های سلسله مراتبی[2] استفاده شده است[۸-۱۰]. از دیگر روش‌های برآورد مدل استفاده از شبکه عصبی چند لایه پرسپترون است که در بسیاری از مقالات از این روش استفاده شده است[۱۱]. همچنین استفاده از برآورد کننده بیزی در آموزش اختلاط خبره ها از دیگر روش های بکار برده شده است[۲].

۱-۴ هوش مصنوعی

هوش مصنوعی قابلیتی است برای انجام فعالیت هاو وظایف رایانه که همانند انسان نیاز به دانش، دقت، استدلال، آموزش، فهم قابلیت های ادراکی دارد.

ریشه اصلی سیستم های خبره[3] و یا سیستم های مبتنی بر دانش[4](KBS) به حوزه مطالعاتی به نام هوش مصنوعی[5](AI) برمی‌گردد. تحقیق درباره‌ی AI در دهه ۱۹۴۰ شروع شد، زمانی که اولین نسل رایانه‌ها در مؤسسات تحقیقاتی به کار برده شد، اساس منطق ماشین مبتنی بر علم ریاضیات بوده است و افرادی مانند گرت گودل، الونزو چاچ، آلوین ترینگ با استفاده از روشهای فرموله شده‌ای در استدلال منطقی به آن پرداخته بودند.

تحقیق اصلی این افراد بر محاسبه پیش‌بینی و قیاس منطقی[6] متمرکز شده است. این نحوه فرموله کردن نقش مهمی در سیستم‌های AI داشته است از پیامدهای توجه به AI در حل مسائل از حوزه ریاضی به حوزه مسائل دنیایی واقعی با تمرکز به فنون حل مسائل کلی معمول از این طریق، ظهور KBS و سیستم های خبره تجاری بوده است .

سیستم های مبتنی بر دانش شامل عملگر هایی است که مشخص می‌کندکه چطور یک سیستم از یک وضعیت می‌تواند به وضعیت بعد ونهایتا به سوی وضعیت هدف پیش رود. برای مثال در یک بازی شطرنج عملگرها قواعدی هستند که مهره های مختلف را قادر به تغییر مکان از یک خانه به خانه دیگر می‌کند.از نقایص مدلهای موجود این بود که با افزایش اندازه مسائل فضای جستجوی آنها به صورت نمایی افزایش می‌یافت،بنابراین محققین با تمرکز بر دانش خاص در هرمساله، توانستند فضای جستجو را محدود تر نمایند. در سال ۱۹۸۶ واترمن،برنامه ای با کیفیت بالا بنحوی که دانش خاصی در حوزه آن مساله مرتبط باشد،را طراحی کرد.اولین فعالیت‌های KBS در حوزه زندگی واقعی نظیر تشخیص عفونی و پیش‌بینی منابع طبیعی معدنی در نواحی مختلف جغرافیایی دنیا بوده است[۱۲].

---
[1] Expectation – maximization algorithm
[2] Hierarchical mixtures of experts
[3] Expert systems
[4] Knowledge-based systems
[5] Artificial intelligence
[6] Logical reasoning

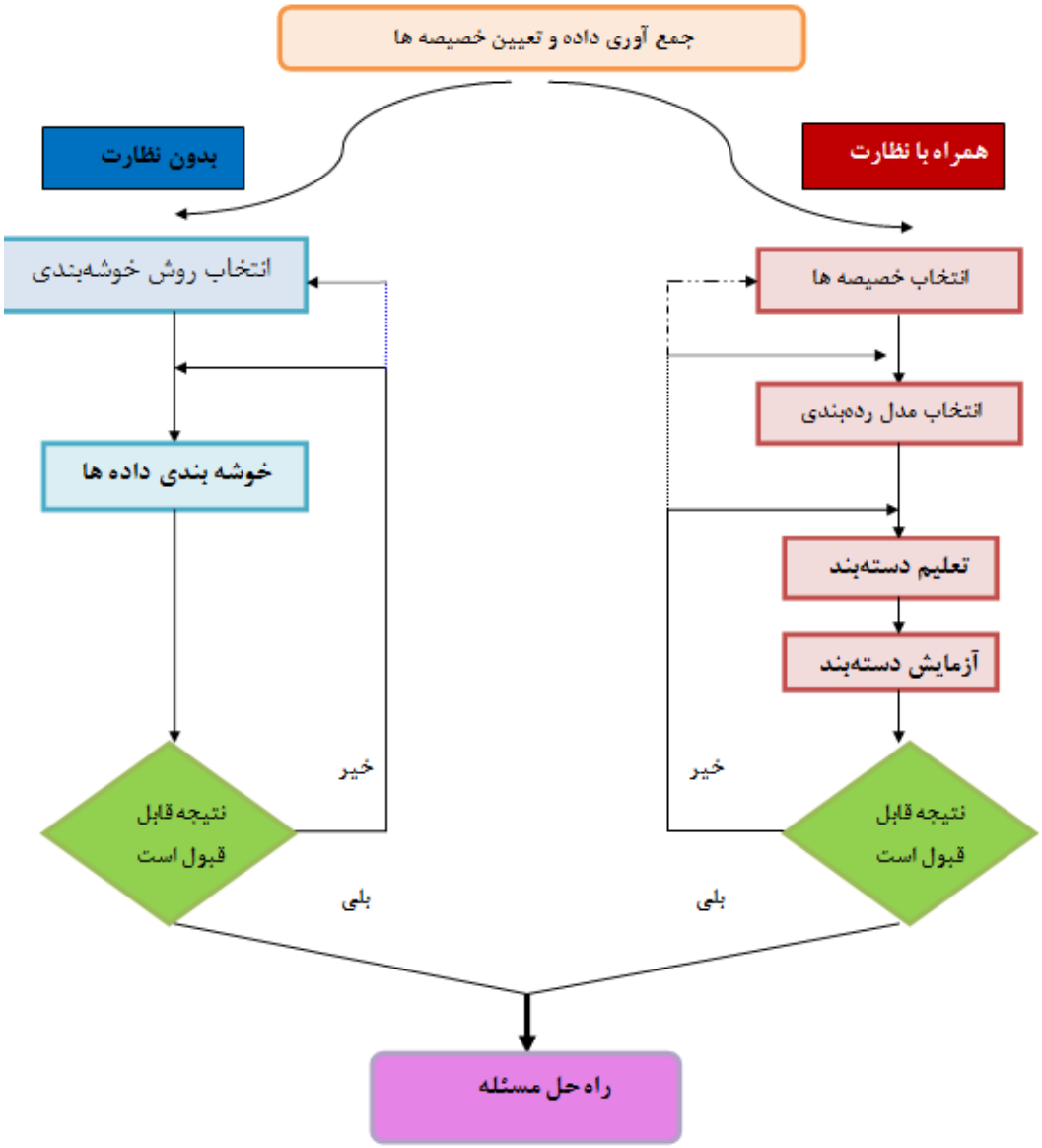
۱-۵ یادگیری ماشین

در سال ۱۹۵۹ اصطلاح یادگیری ماشین اولین بار توسط، ساموئل مطرح شد که به بررسی توانایی آموزش رایانه بدون وجود برنامه نویسی صریح می‌پردازد. به عبارت دیگر یادگیری ماشین راه هایی را که رایانه می‌تواند به طور مستقیم از داده ها دانش به دست‌آورد را بررسی می‌کند .

در سال ۱۹۶۳ مورگان و همکارانش در مقابل فرض های محدود کننده آمار کلاسیک از روش درخت رگرسیونی، که روشی بدون فرض های آماری می‌باشد را معرفی کردند . این روش به عنوان اولین روش غیر آماری یادگیری ماشین در حل مسائل آماری معروفمی‌باشد.این الگوریتم ها با هدف بهبود عملکرد از طریق تجربه عمل می‌کنند. در حدود سالهای ۱۹۸۰ برهم کنش بین دو رشته آمار ویادگیری ماشین شروع شد، محققین یادگیری ماشین با سه مسأله رگرسیون،رده‌بندی، خوشه‌بندی توسط آماردانها آشنا شدند سپس آنها شروع به جستجو راه‌حل های غیر آماری و ناپارامتری جهت حل این مسائل کرده‌اند[۱۳].

یادگیری ماشین یک شاخه مهم از گرایش هوش مصنوعی است ، که ترکیبی از روش های ابتکاری هوش مصنوعی به همراه تحلیل آماری پیشرفته می‌باشد. که هدف از آن تعلیم یک ماشین است بطوریکه بتواند تجربیات و نمونه های موجود را یاد بگیرد. حاصل این یادگیری ایجاد یک مدل طبقه بندی است که بر اساس آن ماشین می‌تواند نمونه هایی را که در آینده می‌بیند و مشابه نمونه های موجود هستند در کلاس مناسب خود قرار دهد[۱۴]. امروزه روشهای بازشناسی الگو[1] ، به عنوان یک شاخه از یادگیری ماشین، کاربردهای فراوانی در زمینه های مختلف علمی و صنعتی پیدا کرده‌اند. در فرآیند باز شناسی الگو ،الگوهای ورودی در کلاسها و دسته‌های از پیش تعیین‌شده دسته‌بندی می-شوند. روند بازشناسی الگو در شکل ۱-۱ نشان داده شده است[۱۵].

[1]Patern Recognition



شکل ۰-۱روند بازشناسی الگوشکل

اولین گام در بازشناسی الگو ، جمع آوری تعداد مناسبی نمونه از الگو های مورد نظر است. این بخش زمان زیادی از فرآیند طراحی سیستم باز شناسی الگو را به خود اختصاص می‌دهدو گاهی اوقات با مشکلاتی همراه است. پس از جمع آوری نمونه های لازم ، باید اقدام به انتخاب نوع ویژگی کرد . انتخاب نوع ویژگی نیازمند دانش اولیه در مورد داده ها است . انتخاب باید بگونه‌ای باشد که ویژگی‌ها ، وجه مشترک الگوهای متعلق به یک کلاس و وجه تمایز الگوهای کلاسهای مختلف باشند. توانمندی ویژگی برای جداسازی نمونه‌های کلاسهای مختلف ،معیار انتخاب ان است.

مرحله بعدی انتخاب خصیصه[1] ها به معنیبدست آوردن خصوصیات الگو هاست، به نحوی که این

[1] Features

خصوصیات وجه مشترک الگو های متعلق به یک کلاس و وجه تمایز الگو های کلاس های مختلف باشند. پس از انتخاب مدل رده‌بندی،پارامترها باید مشخص شوند. روشهای مختلفی برای تعیین این پارامترها ارئه شده‌اند. گروهی از این روشها بر اساس خواص آماری، مانند بردار میانگین و ماتریس کواریانس نمونه های آموزشی هر کلاس در هر دسته‌بند بیز، پارامترهای طبقه‌بندی را تعیین می‌کنند. در مقابل گروه دیگری از روشها، با استفاده از الگوریتم های یادگیری تکراری مانند آموزش یک طبقه بند شبکه عصبی بر روی نمونه های یادگیری، پارامترهای رده‌بندی را تعیین می‌کنند. در این روشها پارامترهای یادگیری در طول فرآیند یادگیری تعیین می‌شوند. در یادگیری بدون ناظر ، پس از جمع آوری داده‌های کافی از الگو های مورد نظر ، یک روش خوشه‌بندی انتخاب می‌شود و بر اساس آن ، داده ها خوشه‌بندی می‌شوند. اگر نتیجه خوشه‌بندی قابل قبول نباشد، با تغییر روش خوشه بندی یا تغییر پارامترهای خوشه‌بندی، کار ادامه پیدا می‌کند تا نتیجه مورد نظر حاصل شود.

یکی از ایده های جالب در یادگیری ماشین تجزیه یک مسئله مشکل به زیر مسئله هایی ساده تر و حل این مسائل با تکنیک های استاندارد است. به این سیستم هافرا تحلیل[1] گفته می‌شود. یکی از زیر شاخه های این سیستم ها ،سیستم های یادگیرنده چندگانه است.که در آن محققین از یادگیرنده های چندگانه برای یک مسئله یکسان، وترکیب نتیجه آنها استفاده می‌کند. گروه یادگیرنده ها را در این مسائل شورا نامیده می‌شود. روشهای متعددی برای یادگیری ماشین و تشکیل شوراها و ترکیب تصمیم‌ها وجود دارد[۱۵].

در بسیاری از مسائل پیش‌بینی‌کننده احتیاج به تصمیم گیری در موقعیت های مختلف دارد مانند راننده ای که برای خبره شدن نیاز به یادگیری رانندگی در جاده ،در بارندگی و جاده های یخ زده دارد. در مسائل رده‌بندی نیز هر یک از دسته‌بندها نقاط قوت و ضعف متفاوت دارند و رقابت و مدیریت این دسته‌بندها در سیستم های شورایی با توجه به درک این محدودیت ها در پیش‌بینی داده ها به دست می‌آید. محدودیت در اینجا به معنای ضعف در پیش بینی است که موجب خطا می‌شود.

۱-۶ پیش پردازش داده‌ها

کیفیت داده ها در استخراج نتایج مطلوب و اطلاعات حقیقی بسیار موثر است ، پایگاه های داده اغلب دارای داده های پرت، گمشده و نا ایستا هستند. بنابراین برای ارتقاء کیفیت داده ها لازم است در ابتدای کار داده ها به صورت زیر پردازش شوند.

پاکسازی داده ها : برای انجام یک تحلیل مطلوب لازم است مقادیر گمشده جایگزین شوند ، و داده های مزاحم شناسایی و به نحوی مناسب با آنها برخورد و نا ایستایی ها اصلاح شوند.

الف) مقادیر گمشده وجود مقادیر گمشده در داده ها می تواند تحلیل داده ها را بسیار دشوارسازد. در صورت وجود مقادیر گمشده در داده ها باید به گونه ای مناسب در مورد آنها تصمیم

[1]Meta Analysis