به نام آنکه

جان را

فکرت

آموخت

تقدیم به درگاه مطهرش؛

کاش این تلاش ناچیز و این عمر رفته، اندکی روح تاریک و خطاکارم را جلا بخشیده باشد.

دانشگاه صنعتی امیرکبیر

دانشکده مهندسی برق

پایان‌نامه کارشناسی ارشد

مهندسی برق- مخابرات سیستم

طبقه‌بندی کلمات متن تایپی فارسی-انگلیسی بر حسب نوع قلم، در قالب نظریه بیزی، و با استفاده از ویژگی‌های فیلتر گابور

نگارش

موناحاجی‌مومنی

استاد راهنما

دکترحمیدرضا امین‌داور

استاد مشاور

دکتر کریم فائز

بهمن ماه۸۷

تقدیم به خانواده صبورم،

پدر مهربانم، مادر فداکارم و خواهر زیباشناسم سها، که هنر را دوست دارد؛

تقدیم به دوست مهربان و خونگرمم نسترن، که می‌خواهم همیشه فارغ از غصه باشد، آنگونه که هست؛

تقدیم به دوستان عزیزم لیلا، حمیده، رویا، زینب، و هدی که صحبت‌های دلنشینشان بر جانم می‌نشست، همواره؛

تقدیم به خاله، شوهر خاله و خاله زاده‌های پر کوششم سارا، سینا، سعید که همیشه در کنارم بوده‌اند؛

تقدیم به کوشان، پسردایی جوانم که دانش‌اش همیشه تحسینم را بر می‌انگیزد؛

تقدیم به تمام اساتید زبانم و هم‌کلاسی‌های با نشاطم در موسسه کیش، که بسیار دلتنگشان شدم؛

تقدیم به تمام هم‌کلاسی های محجوب و خداپرستم در کلاس سنتور؛

تقدیم به همکاران محترمم در شرکت کام ارتباطات که بی‌شک از مشغله فراوان من آزرده شدند؛

تفدیم به حضرت حافظ که «دل رمیده ما را رفیق و مونس شد» ؛

تقدیم به ساز پاکم، سنتور :

مطرب کجاست تا همه محصول زهد و علم در کار چـــنگ و بربط و آواز نی کنم

از قـیـل و قـال مـدرسه حالـی دلـم گرفت یکچند نیز خدمت معـشوق و می کنم

تقدیم به خوشنویسی، نقاشی، به شعر و به تمام کتاب‌های نخوانده، که دلم برایشان پر می‌کشد؛

تقدیم به طبیعت شگفت، هر آنچه دیده و یا ندیده‌ام؛

تقدیم به نظام عظیم هستی، قوانین بی پایان ولی متحدش و تمام دانشجویانش؛

.. ..

.. ..

سپاس‌گزاری

«منت خدای را عزّ و جل که طاعتش موجب قربت است و به شکر اندرش مزید نعمت. هر نفسی که فرو می‌رود ممد حیات است، و چون برمی‌آید مفرح ذات. پس در هر نفسی دو نعمت موجود است، و بر هر نعمتی شکری واجب.

از دست و زبان که برآید، کز عهدهٔ شکرش به درآید.

باران رحمت بی‌حسابش همه را رسیده، و خوان نعمت بی‌دریغش همه‌جا کشیده، پردهٔ ناموس بندگان به گناه فاحش ندرد، و وظیفهٔ روزی به خطای منکر نبرد.

فراش باد صبا را گفنه تا فرش زمردین بگسترد، و دایهٔ ابر بهاری را فرموده تا بنات نبات در مهد زمین بپرورد. درختان را به خلعت نوروزی قبای سبز ورق در برگرفته، و اطفال شاخ را به قدوم موسم ربیع کلاه شکوفه بر سر نهاده، عصارهٔ نالی به قدرت او شهد فایق شده، و تخم خرمایی به تربیتش نخل باسق گشته.»

گویند که: «دزدی به خانهٔ پارسایی درآمد؛ چندانکه جست چیزی نیافت؛ دل‌تنگ شد. پارسا با خبر شد، گلیمی که بر آن خفته بود در راه دزد انداخت تا محروم نشود.»

خدای بزرگ و مهربان را به خاطر فرصت‌هایی که در دوران تحصیلم فراهم آمد، شکر می‌کنم. بی‌شک آشنایی با تنی چند از اساتید بزرگوار و یافتن دوستانی عاقل و شاد، با ارزش‌ترینِ هدیه‌ها بودند.

سپاس و تقدیر فراوان، شایستهٔ استاد محترمم جناب آقای دکتر امین‌داور است، که گوشه‌ای از دانش گرانبهایشان را به من بخشیدند. شکر و حمد سعدی، و حکایت نغز او، بهتر و بیش‌تر گویای این احوال است. آری، این است اخلاق بزرگان.

از جناب آقای دکتر آقایی‌نیا، که دلسوزانه پایه‌های منطقی و فلسفی این دوره را در ذهنم برپا کردند، بسیار ممنونم.

از جناب آقای دکتر کلانتری، عضو هیئت علمی دانشگاه صنعتی خواجه نصیر، که سخت-گیرانه در امر آموزش کوشیدند، روشنگری و واقع‌بینی خاص‌شان را با پیام مقاومت مدوله کردند و به گوش امثال من رسانیدند، تشکر فراوان دارم.

این سه بزرگ، به تمامی و با شیوهٔ خویش ندای حق‌پرستی را در من زنده نگاه داشتند. خوشبخت بودم به سبب دیدارشان.

از جناب آقای دکتر فائز که با گشاده‌دستی وقت گرانبهای خویش را صرف راهنمایی و هدایت اینجانب ـ به‌ویژه در بخش شناسایی الگو و جمع‌آوری داده ـ نمودند، کمال قدردانی را به جا می‌آورم.

در آخرین روزهای نگارش نسخهٔ اولیهٔ این پایان‌نامه از حضور اساتید محترم ارزیاب نیز بهره‌مند گشتم. از جمله، جناب آقای دکتر مقدم‌جو، که با خوش‌خلقی و فروتنی، امیدبخش من در هفتهٔ آخرِ پیش از دفاع بودند، و نیز جناب آقای دکتر بخشی، عضو هیئت علمی دانشگاه شاهد، که با دیدگاه‌های ریزبینانه و واقع‌گرای خودشان، مرا به مسئولیت خویش در قبال موضوع تحقیق و نگارش آن، بیش از پیش واقف ساختند. هیچ‌چیز نشاط‌آورتر از آن نیست که انسان را لایق تحقیق شمارند، و رسالتی را بر عهدهٔ او بنهند.

بی‌شک اما، بدون خانوادهٔ مهربانم، مجال آشنایی عزیزانی چنین را نمی‌یافتم. درود و سپاس به پیشگاهشان، که مرا در این راه هدایت و تشویق کردند.

صحبت از خوبی پایانی ندارد، اما، خوشحالم که موهبت تحصیل نصیبم شد، و خوشحالم که پایان این کار در زمستان بود و هماهنگ با نظام طبیعت. باشد که فصل بهار و عید نوروز، جوانه‌های این موهبت را شکوفا کند، تا این بار ما بخشنده باشیم، به امید خدا.

«حافظ این حال عجب با که توان گفت که ما بلبلانیم که در موسم گل خاموشیم».

نوشته شد در بهار ۸۸ ، پایان.

پیش‌درآمد

طبقه‌بندی کلمات در متون تایپی برحسب نوع قلم، کاربردهای فراوانی در پردازش اتوماتیک اسناد و نوشته‌ها دارد که در دنیای امروز هم رایج شده و هم ضروری به نظر می‌رسد. در این میان، نفس عمل طبقه‌بندی از هر نوع که باشد در افزایش دقت و بهبود عملکرد کل مسأله به شدت مؤثر خواهد بود. مثالی از این دست می‌تواند طبقه‌بندی کلمات یک صفحه از لغت‌نامه‌ای باشد که هر نوع قلم، اطلاعات جانبی مختلفی را در مورد کلمهٔ مذکور نمایان سازد: قلم ضخیم نشانهٔ شروع یک مدخل جدید باشد و قلم مایل جایگاه دستور زبانی کلمه. قلم عادی با فونت کوچک اطلاحات را مشخص می‌کند و قلم عادی با فونت بزرگتر، معنی را. آنچه در این پروژه بدان پرداخته شده است، پیاده سازی الگوریتمی در طبقه‌بندی کلمات یک متن تایپی، برحسب نوع قلم نگارش آن است. بنابراین فرض می‌شود که کلمات متن مورد نظر با یکی از چهار نوع قلم عادی، مایل، ضخیم، و مایل ضخیم، نوشته شده‌اند و هدف، طبقه‌بندی آنهاست.

از دیدگاه کلی، هدف پروژه، یافتن یک توزیع احتمال است که پارامترهای آن، توصیف کنندهٔ کلاس‌های مورد نظر در مسألهٔ فیزیکی‌اند. کل روند تحقیق در سه بخش قابل پردازش است. بخش اول، توصیف فیزیک مسأله به فضای ریاضی ویژگی‌هاست. بخش دوم، انتخاب توزیع احتمالی برای مدل کردن فضای به دست آمده است، و بالاخره بخش سوم، به کارگیری روشی برای تخمین پارامترهای مدل احتمال مذکور. پیداست که از نظر موضوعی، در دستهٔ تخمین مدل‌های احتمال پارامتری جای داریم. به این موضوع، قالب بیزی هم اضافه می‌شود، و پارامترهای مدل احتمال، متغیرهای تصادفی با توزیع احتمال پیشین فرض می‌شوند. با این زمینه، تکیهٔ اصلی بر الگوریتمی تکرارشونده در زمینهٔ تقریب‌های توزیعی است که همراه با یک مدل توزیع احتمال جدید، زیربنای تحلیلی پروژه را تشکیل می‌دهند. منظری که از ابتدا تا انتهای این نوشته، سعی در توصیف آن داریم، در واقع ارتباط این دو مورد است، «*تا که قبول افتد و که در نظر آید*».

کلمات کلیدی: فیلتر گابور(*Gabor Filter*)، ویژگی(*Feature*)، مدل ترکیبی(*Mixture Model*)، تحلیل بیزی(*Bayesian Analysis*)، تقریب تغییری (*Variational Approximation*)، توزیع دیریکلهٔ تعمیم-یافته(*Generalized Dirichlet Distribution*).

راهنمای نشانه‌ها و برخی نام‌گذاری‌ها

در این پایان‌نامه به منظور سهولت ارجاع و روشن‌تر شدن روابط موجود، برخی اسامی و قراردادها در تمام فصول رعایت شده است، که در زیر می‌آید. بدیهی است سایر اسامی متغیرها، که ممکن است برای توصیف چند مفهوم مختلف و در فصل‌های غیریکسان به‌کار رفته باشند، در متن موجود به فراخور کاربرد، تعریف خاص به خود را خواهند داشت. در این نام‌گذاری، بردارها با قلم ضخیم، و مجموعه‌ای از چند بردار، به صورت قلم ضخیم و یک نشانهٔ بردار بالای آن نوشته شده است.

ابعاد بردار ویژگی	d
تعداد مؤلفه‌های مدل ترکیبی	M
تعداد نقاط داده (مشاهدات)	N
ویژگی k ام از بردار داده	X_k
بردار ویژگی(داده/مشاهده)	$\mathbf{X} = (X_1, X_2, ..., X_d)$
مجموعه‌ای از بردارهای ویژگی	$\underset{\sim}{\mathbf{X}} = \{\mathbf{X_1}, \mathbf{X_2}, ..., \mathbf{X_N}\}$
متغیر نهان یک نقطهٔ داده در مدل ترکیبی	$\mathbf{z}$
مجموعه‌ای از N متغیر نهان برای N مشاهده	$\underset{\sim}{\mathbf{z}}$
واریانس	v یا Var
میانگین	m

بردار پارامتر یک مؤلفه	$\boldsymbol{\theta}$
مجموعهٔ N بردار پارامتر	$\boldsymbol{\Theta}$
تابع توزیع احتمال	$p(.)$
عضوی از مجموعهٔ توابع توزیع احتمال قابل دستگیری در قالب تقریب تغییری بیزی	$\hat{p}(.)$
تقریبی از توزیع احتمال واقعی	$\tilde{p}(.)$
تابع درست‌نمایی	$L(.)$
تابع دی‌گاما	$\Upsilon(.)$
فاصلهٔ لگاریتمی *Kullback-Leibler*	$KL(.)$
ضرایب شرکت مؤلفه‌ها در مدل ترکیبی	$P(j);\ j = 1,...,M$
بردار شامل تمامی ضرایب شرکت مؤلفه‌ها	$\boldsymbol{\mu} = (P(1),...,P(M))$
تمام بردارها سطری هستند و ضرب داخلی آنها به صورت $\mathbf{AB}^T$ تعریف می‌شود، که در آن، $\mathbf{A}$ و $\mathbf{B}$، بردارهای هم اندازه، و $\mathbf{B}^T$ ترانهادهٔ ماتریس $\mathbf{B}$ است.	

فهرست

فصل اول : درآمدی بر موضوع ۱

۱-۱- گفتاری کلی پیرامون مدل فیزیکی مسأله ۲

۱-۲- درآمدی بر نظریهٔ بیزی ۵

۱-۳- مختصری بر پیشینهٔ تاریخی ۷

۱-۴- درآمدی بر تقریب توزیعی تغییری در چارچوب بیزی ۱۴

۱-۴-۱- انتخاب روش تقریب توزیعی از دیدگاه بیزی ۱۴

۱-۴-۲- مروری اجمالی بر تقریب‌های توزیعی ۱۷

۱-۴-۳- بیان رسمی قضیهٔ تقریب *VB* ۱۸

۱-۵- خلاصه ۲۲

۱-۶- دورنمای فصل‌های آینده ۲۲

فصل دوم : طرح مسأله و توصیف ریاضی ۲۳

۲-۱- طرح موضوع ۲۴

۲-۲- فیلتر گابور ۲۶

۲-۲-۱- طراحی بانکی از فیلترهای گابور ۲۸

۲-۲-۲- استخراج ویژگی‌ها به کمک فیلتر گابور ۳۰

۲-۳- توزیع دیریکلهٔ تعمیم یافته ۳۳

۲-۴- مدل ترکیبی توزیع‌های دیریکلهٔ تعمیم یافته ۳۶

۲-۵- الگوریتم تکرارشوندهٔ *IVB*..... ۳۸

۲-۵-۱- مراحل پیاده‌سازی الگوریتم *IVB* ۳۹

۲-۵-۲- معرفی توزیع احتمال‌های خانوادهٔ نمایی..... ۴۵

۲-۵-۳- توزیع احتمال خانوادهٔ نمایی و تقریب *VB*..... ۴۷

۲-۵-۴- بازنگری توزیع دیریکلهٔ تعمیم یافته به عنوان عضوی از خانوادهٔ نمایی..... ۵۰

۲-۵-۵- اعمال مراحل هشت‌گانهٔ تقریب *VB* به توزیع دیریکلهٔ تعمیم یافته ۵۳

۲-۶- بلوک طبقه‌بندی و تصمیم‌گیری ۶۱

۲-۷- خلاصه ۶۳

فصل سوم: تحلیل و بازنگری..... ۶۵

۳-۱- نقش دو جانبهٔ فیلتر گابور، ایده‌ای در انتخاب ویژگی‌ها..... ۶۶

۳-۲- مقداردهی اولیهٔ الگوریتم *IVB* بر مبنای مجموعهٔ نقاط داده..... ۷۰

۳-۲-۱- بررسی تناسب توزیع داده با توزیع دیریکلهٔ تعمیم یافته..... ۷۰

۳-۲-۲- شرحی بر روش مقداردهی اولیهٔ الگوریتم *IVB*..... ۷۳

۳-۳- الگوریتم *EM* در تخمین *MAP*..... ۸۰

۳-۳-۱- استخراج *EM* از روش تقریب توزیعی *VB*..... ۸۰

۳-۳-۲- تحقق مرحلهٔ ماکسیمم‌سازی *EM* به روش فیشر..... ۸۵

۳-۴- خلاصه..... ۸۸

فصل چهارم: نتایج حاصل از شبیه‌سازی..... ۸۹

۴-۱- نتایج استخراج ویژگی‌ها از داده‌های آموزش..... ۹۰

۴-۱-۱- روش جمع‌آوری داده..... ۹۰

۴-۱-۲- نمایش تناسب ویژگی‌های استخراج شده از نقاط داده با توزیع دیریکلهٔ تعمیم یافته..... ۹۲

۴-۲- نمایش نقاطی از فضای پارامترهای توزیع دیریکلهٔ تعمیم یافته..... ۹۳

۴-۳- اجرای الگوریتم *IVB* و مقداردهی اولیهٔ ابرپارامترها..... ۹۶

۴-۴- اجرای طبقه‌بندی براساس ماکسیمم احتمال شرطی کلاس‌ها..... ۹۷

۴-۵- خلاصه..... ۹۹

بحث و نتیجه‌گیری..... ۱۰۰

پیشنهادات ۱۰۲

مراجع..... ۱۰۳

فصل اول

درآمدی

بر

موضوع

۱-۱- گفتاری کلی پیرامون مدل فیزیکی مسأله

مسأله طبقه‌بندی کلمات بر حسب نوع قلم در مدارک و اسناد تایپیِ ساختاریافته، از اهمیت ویژه‌ای برخوردار است، چرا که بدونّ کشفّ نظمِ حاکم بر چنین نوشته‌هایی، رعایتّ اصول، ساختار و سازمان-یافتگی در آن, امری عبث به نظر می‌رسد. اگرچه قدرت درک پیچیده و خیره‌کنندهٔ انسان از طریق ارتباط توأم و آنی چشم و مغز، این ساخت‌یافتگی را تشخیص می‌دهد، انتقالِ چنین مفهومی به ماشین، که فاقد هوش و حواس است، می‌تواند بسیار دشوار باشد. بیان دیگر آنکه ساختار مدارک و نوشته‌های تایپی به زبان تصویرحاوی اطلاعاتی برای بیننده (ویا پردازشگر) است که به منظور برخورداری از آن می‌بایست کد ساختاری نوشته گشوده شود. این همان هدفی است که طبقه‌بندی دنبال می‌کند، یعنی شناسایی آنکه یک کلمهٔ خاص چه کد ساختاری دارد. این کد ساختاری همان نوع قلم است که در اینجا مبنای طبقه-بندی قرار می‌گیرد.

نوع قلم در این پروژه به چهار نوع عادی، مایل، ضخیم و مایلِ ضخیم است. بنابراین از دیدگاه شناسایی آماری الگو، ما دارای چهار کلاس هستیم و هر کلمه قرار است در یکی از این چهار کلاس جای بگیرد. به‌طور کلی، تمام روش‌های طبقه‌بندی درحالت ایده‌آل و مینیمم خطا، با قاعده بیز توصیف می‌شوند، اما برحسب نوع مسأله، طبقه‌بندی کننده می‌تواند به حالت‌های ساده‌ترِ خطی، می‌نیمم فاصله و مانند آن، تبدیل شود. در استخراج رابطهٔ طبقه‌بندی بیز که خود تابعی از توزیع احتمال داده و احتمال کلاس‌هاست، نیاز به یک فرآیند یادگیری است. این یادگیری می‌تواند در حضور ناظر (*Supervised*)، بدون ناظر (*Unsupervised*)، و یا بخشی در حضور ناظر (*Semi-supervised*)، باشد. در هر یک از این انواع، فرآیند یادگیری با پردازش بخشی از اطلاعات، موفق به طراحی پارامترهای مدل توزیع احتمال داده‌ها و در نتیجه تابع طبقه‌بندی می‌شود. آنچه به عنوان اطلاعات نام بردیم، در این پروژه همان تصویرِ کلمات متن است، با این تغییر که باید به نحوی تصویر کلمه را برای ماشین (کامپیوتر) قابل فهم نمود. به این جهت از هر تصویرِ کلمه، تعداد مشخصی ویژگی (*Feature*) استخراج می‌شود که به صورت برداری و متناظر با تصویر مربوطه‌اش ذخیره می‌شود. مجموعه تمام این بردارها فضای ویژگی را خلق می‌کند که ابعاد آن برابر با تعداد ویژگی‌های استخراج شده است. فضای ویژگی ارتباط مستقیمی با نوع ویژگی‌هایی

که از تصویر استخراج می‌شود دارد و روشن است، که هرچه پیچیدگی این فضا کمتر باشد، تابع طبقه بندی اقبال بیشتری در جداسازی داده‌ها از یکدیگر دارد. از این دیدگاه، بخش استخراج ویژگی از داده‌ها را می‌توان زیرمجموعه‌ای از بخش طبقه‌بندی دانست و تلاش هر طراح باید آن باشد که ویژگی‌های استخراج شده هم‌جهت و هم‌هدف با بخشِ طبقه‌بندی باشند.

استخراج ویژگی از تصاویر به روش‌های گوناگونی قابل انجام است و آنچه که حائز اهمیت است، انتخاب روشی است که بهترین توصیف را از داده‌ها با در نظر گرفتن هدف طبقه‌بندی به دست می‌دهد. ویژگی‌هایی که می‌توانند در مسألهٔ ما قلم‌های مختلف را از هم متمایز سازند، زاویه از خط افق، نسبت چگالی نقاط سیاه به سفید، نسبت پهنای کلمه در جهت عمودی به افقی و ... است. استخراج این نوع ویژگی‌ها به روش‌های پردازش تصویر ممکن و میسر بوده و در زمینه‌های گوناگون مورد استفاده واقع شده است. به جز آن، ویژگی‌های دیگری که بیش‌تر از خواص آماری تصاویر بهره می‌برند، چون گشتاورهای هندسی دوبعدی که تا مرتبه چهارم مورد استفاده قرار گرفته‌اند، و یا گشتاورهای زرنیک (*Zernike*)، و البته در سال‌های اخیر استفادهٔ به‌جا از ضرایب موجک با موجک‌های مادرِ متفاوت، همگی از عناوین مطرح در زمینه استخراج ویژگی از تصاویر هستند. گشتاورهای هندسی را می‌توان متناظر با ماسک‌هایی دانست که هریک تصویر ورودی را به نحوِ خاصی فیلتر می‌کنند. گشتاورهای زرنیک نیز تا کنون در کاربردهایی چون تشخیصِ ارقامِ صفر تا نُه به صورت دست‌خط مورد استفاده قرار گرفته‌اند. ضرایب موجک بیشتر جهت تفکیک تصویر به باندهای فرکانسی مختلف کاربرد دارند و علاوه بر آن، تصویر را در سه جهت عمودی، افقی و قطری تجزیه می‌کنند. قبل از بحثّ موجک‌ها، دنیس گابور تئوری خود را در مورد سیگنال‌هایی که هم در زمان و هم در فرکانس محدودند، ارائه کرده بود و موجک‌های گابور (*Gabor*) هم پس از آن مطرح شدند. اما جدا از نظریه موجک‌های گابور، که شاید با ظهور انواع امروزی و متنوع موجک‌ها همچون *Edgelet*، *Contourlet* و ... کمتر مورد استفاده قرار بگیرند، کاربرد فیلتر گابور در شناسایی الگوها و بافت‌ها هنوز جایگاه خود را حفظ نموده است. این امر نتیجه‌ای از شباهت موجود در ساختار این فیلتر و ساختار سلول های بینایی مغز است، چرا که به اثبات رسیده است که مغز انسان در تشخیص بافت‌ها و الگوهای تکرارشونده، از دو کانال مجزا با اختلاف فاز نود درجه بهره می‌جوید که بسیار منطبق با کانال‌های زوج و فرد فیلتر گابور است. اگرچه مسألهٔ طبقه-

بندی کلمات برحسب نوع قلم، چندان با مدلِ طبقه‌بندی بافت‌ها که عموما شکلِ متناوب و تکرارشونده دارند، سازگار نیست، از ویژگی‌های استخراج شده توسط گابور تاکنون در زمینهٔ تشخیص و شناسایی فونت‌های متونّ انگلیسی، چینی و...، و نیز، فونت‌های متون فارسی استفاده شده است. البته مسأله تشخیص فونتّ یک متن، با تشخیص قلمِ یک کلمه خاص در کل متن، نسبتاً متفاوت است و متن نوشته شده با یک نوع فونت، بیش‌تر به مدل بافت‌های تکرارشونده نزدیک است. با این وجود برخی از محققان با تکرارِ کلمه در پنجره‌ای به طول ثابت سعی کرده‌اند شرایط بهره‌مندی از ویژگی‌های گابور (*Gabor Features*) را در کار خود را فراهم کنند. فیلتر گابور انعطاف زیادی در تعیین جهت و فرکانس مکانی به منظور ورود به بخش‌های جزیی یا کلی تصویر و نیز پهنای باند عبور فرکانس‌ها حول فرکانس مرکزی دارد. خاطرنشان می‌شود که آنچه در جداسازی نوع قلم مفید است استخراج ویژگی‌های کلی از تصویرِ کلمه است و اگر به دنبال ویژگی‌های نهفته در جزئیات برویم بیشتر به ساختار کلمه نزدیک شده‌ایم تا ساختارِ متن. ساختارِ کلمه به شدت از دیدگاه هدفی که در این پروژه داریم دردسرساز است. به عنوان نمونه در [۱]، طبقه‌بندی کلمات انگلیسی تایپی برحسبِ نوع قلم با این فرض انجام شده است که نتایج تشخیص حروف (*OCR*[۱]) انگلیسی موجود است و طبقه‌بندی از سطح حرف‌های یکسان انگلیسی آغاز شده و نتیجه بیست و شش طبقه‌بندی روی تمام حروف به منظورِ طبقه بندی کلمات به کار گرفته شده است. به این ترتیب تشخیص نوع قلم روی زیرمجموعهٔ حروفی که تصویر و شکل هندسی یکسانی دارند، انجام می‌گیرد. در مورد زبان فارسی، این کار هم از پیچیدگی و هزینه بیشتر و هم از دقت پایین‌تری برخوردار است، چرا که حروفّ فارسی به هم چسبیده‌اند و هر یک می‌توانند چندین شکل داشته باشند. علاوه بر آن، منطق یاد شده به سختی قابل پذیرش است، چرا که طبقه‌بندیِ کلمات برحسب نوع قلم برای انسان، از تفکیک حرف به حرف راحت‌تر است و نمی‌توان فرض کرد که در دورنمای کلی مسأله دانشی نسبت به حروف تشکیل‌دهندهٔ کلمات فارسی موجود باشد ولی نسبت به نوعِ قلمی که کلِ کلمه با آن نوشته شده موجود نباشد. تأکید می‌شود که این کار ممکن است، به‌ویژه به

---

[۱] - واژه OCR مخفف Optical Character Recognition است، و امروزه به جز معنای علمی، یک نام تجاری نیز برای اشاره به کاربردهای تشخیص و طبقه‌بندی اتوماتیک و ماشینی اسناد، به شمار می‌رود. در اینجا و هر جای دیگر از این مقال، همواره معنی تجاری آن مورد نظر است.

مدد نرم‌افزارهای تشخیص حروف فارسی که امروزه دقت قابل‌قبولی نیز دارند، اما از واقعیت مسأله در این پروژه به نظر دور می‌رسد.

در ادامهٔ این بخش، ابتدا مختصری از معرفی روش تخمین بیزی (*Bayesian*) آورده خواهد شد و در نتیجهٔ مطالب گفته شده به تدریج به یک دید تحلیلی از روش‌هایی که در این پروژه مورد استفاده قرار خواهند گرفت دست می‌یابیم، که زمینه را برای ورود به جزءییات و برداشت‌های عمیق‌تر از موضوع در فصل‌های ۲ و ۳ هموارتر می‌سازد. در پایان نیز به اختصار مروری بر پیشینهٔ تاریخی تحقیقات برجسته در زمینه‌های مرتبط با این پروژه خواهیم داشت.

۱-۲- درآمدی بر نظریهٔ بیزی

در قالب بیزی که در این پروژه مورد استفاده قرار گرفته است، فرض می‌شود که هر پارامتر یک متغیر تصادفی است و بنابراین دارای یک توزیع احتمال با ابرپارامترهای مشخص است. در واقع قانون بیز که توزیع پسین پارامترها را برحسب مدل مشاهدات و توزیع پیشین پارامترهای مذکور به دست می‌دهد، منطقی‌ترین راه شناختّ ناشناخته‌هاست [۲]. از این منظر، ذهنیت و تمام ارزیابی‌های کمّی ما نسبت به مسألهٔ فیزیکی به‌وسیلهٔ یک مدل احتمال مشاهدهٔ پارامتری به صورت $p(\mathbf{X}|\boldsymbol{\theta})$ توصیف می‌شود و به این‌ترتیب آگاهی از $\boldsymbol{\theta}$، ذهنیت ما نسبت به مشاهدهٔ $\mathbf{X}$ را تعیین می‌کند و نه خود آن را. در عمل اما نتیجهٔ مشاهدات ما از یک آزمایش آن است که مجموعهٔ داده‌های مشاهده، $\underset{\sim}{\mathbf{X}} = \{\mathbf{X}_n, \quad 1 \leq n \leq N\}$، را داریم و می‌خواهیم از پردازش آن به سیستمی که این داده‌ها را تولید کرده است و با مجموعهٔ پارامترهای مجهول q مشخص می‌شود، برسیم. این نحوهٔ یادگیری، به کمک تشکیل مدل احتمال $p(\boldsymbol{\theta}|\underset{\sim}{\mathbf{X}})$، که نمایندهٔ ذهنیت ما نسبت به مدل فیزیکی پس از مشاهدهٔ داده است و به اختصار مدل مشاهدهٔ پسین خوانده می‌شود، شناسایی سیستم را انجام می‌دهد. بیان این مسألهٔ معکوس به کمک قانون بیز چنین است:

$$p(\boldsymbol{\theta}|\underset{\sim}{\mathbf{X}}) \propto p(\underset{\sim}{\mathbf{X}}|\boldsymbol{\theta})p(\boldsymbol{\theta}).$$

(۱-۱)

که نشان می‌دهد چگونه ذهنیت پیشین ما که با مدل احتمال $p(\boldsymbol{\theta})$ توصیف می‌شود، به کمک مشاهدات $\underline{\mathbf{X}}$ به روز می‌شود.

از آنجا که در نظریهٔ بیزی تمام تصمیم‌گیری‌ها و آزمون‌ها بر اساس قوانین احتمال و مدل‌های آن انجام می‌شود، به وفور با انتگرال‌هایی از توابع احتمالِ وزن‌دار برحسب پارامترها و توزیع پسین آنها، $p(q|\underline{\mathbf{X}})$، روبرو خواهیم بود. به عنوان مثال، ممکن است علاقه‌مند باشیم توزیع مشاهدات را برای دادهٔ جدید $\mathbf{X}\not\in$، پیش‌بینی کنیم:

$$p(\mathbf{X}\not\in|\underline{\mathbf{X}}) = \int p(\mathbf{X}\not\in|\boldsymbol{\theta})\,p(\boldsymbol{\theta}|\underline{\mathbf{X}})d\boldsymbol{\theta}. \quad (۲-۱)$$

اما در بسیاری از موارد محاسبهٔ توزیع پسین برای تمام پارامترها دشوار و پیچیده است. به این لحاظ ممکن است بخواهیم یک تخمین نقطه‌ای از آن را بدست آوریم. در این‌صورت یک انتخاب منطقی، محتمل‌ترین مقدار پارامتر به شرط دادهٔ مشاهده شده خواهد بود، که همان تخمین حداکثر احتمال پسین (*Maximum a Posteriori: MAP*) است:

$$\hat{\boldsymbol{\theta}}_{map} = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\underline{\mathbf{X}}) = \arg\max_{\boldsymbol{\theta}} \Big\{\log p(\boldsymbol{\theta}) + \sum_n \log p(\mathbf{X}_n|\boldsymbol{\theta})\Big\}. \quad (۳-۱)$$

انتخاب منطقی دیگر، تخمین حداکثر درست نمایی (*Maximum Likelihood: ML*) است که از دیدگاه فراوانی نسبی و احتمال وقوع مشاهدات قابل پذیرش است، و نیازی به تصمیم‌گیری در مورد توزیع پیشین پارامترها هم ندارد:

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta}} p(\mathbf{X} \mid \boldsymbol{\theta}) = \arg\max_{\boldsymbol{\theta}} \sum_{n} \log p(\mathbf{X}_n \mid \boldsymbol{\theta}). \qquad (۴-۱)$$

اما تخمین حداکثر درست‌نمایی غالبا با این مشکل مواجه است که مدل‌های پیچیده‌تر به طور کلی بیشینهٔ تابعِ درست نمایی بالاتری دارند[۱] و یا آنکه ممکن است تابع درست‌نمایی محدود نباشد. راهکار اصلی در چنین مواردی، اضافه کردن یک جملهٔ کنترل‌کننده[۲] به لگاریتم تابع درست نمایی است. اگر این جملهٔ کنترل‌کننده را به عنوان لگاریتم یک توزیع پیشین تعبیر کنیم، بیشینه‌سازی تابع درست نمایی در این حالت بسیار شبیه به بیشینه‌سازی توزیع پسین خواهد شد. با این وجود، بین این دو شیوه تفاوت‌هایی وجود دارد. یک تفاوت در اینجاست که تخمین *MAP* نسبت به تعریف پارامترها حساس است، اما تابع درست‌نمایی کنترل‌شده این حساسیت را ندارد. در واقع تابع درست نمایی کنترل‌شده، منحصراً یک تابع است، و نه یک چگالی احتمال و به همین دلیل، نسبت به ژاکوبینِ تعریف مجدد پارامترها افزایش و یا کاهش نمی‌یابد[۲].

بنابراین سهولت انتگرالگیری در چنین مواردی، خود مسألهٔ جدیدی است و استفاده از الگوریتم-های بیزی در زمینهٔ پردازش سیگنال بسیار وابسته به و متأثر از روش‌های محاسبه و یا تقریب چنین انتگرال‌هایی است. در مقابلِ استفاده از روش‌های تخمین نقطه‌ای که در پاراگراف قبل مورد اشاره واقع شد، روش‌هایی برای تقریب توزیع احتمال پسین وجود دارد، که در بخش‌های آینده به آن خواهیم پرداخت.

---

[۱] – به این مشکل در اصطلاح Over fitting گفته می‌شود.

[۲] – Regulation Term; Penalized Term

۱-۳- مختصری از پیشینهٔ تاریخی

اکنون که در مورد بخش فیزیکی مسأله و نیز قالب تخمین بِیزی دیدگاه مناسبی ارائه شد، به شرح ریزتر ساختار کل پروژه می‌پردازیم. ماهیت این پروژه چند قسمتی است، به این معنا که از بخش جمع‌آوری داده و انتخاب یک مدل آماری برای توصیف آن، تا بخش طبقه‌بندی و تخمین پارامترهای مدل طبقه-بندی کننده، مورد مطالعه قرار گرفته است. از این جهت، مروری که در این فصل بر روی مقالات و الگوریتم‌های پیشین خواهیم داشت، چند بخش با موضوعیت متفاوت را شامل می‌شود، و چنانچه در انتها نتیجه‌گیری خواهد شد، هدف اصلی این پروژه برقرار کردن ارتباط بین دو ایدهٔ اصلی است. در واقع، بحث تئوری پروژه در قالب روش‌های آماری بِیزی است و آنچه به عنوان ایدهٔ مرکزی به شمار می‌رود، الگوریتمی تکرارشونده به نام *IVB* (*Iterative Variational Bayes*) است، که معادلی برای الگوریتم شناخته‌شدهٔ *EM* در نظریهٔ تخمین کلاسیک به شمار می‌رود. در مقایسه با *EM*، که تخمین نقطه-ای از پارامترهای مدل ارائه می‌دهد، الگوریتم *IVB*، به سیاق سایر روش‌های بِیزی، توزیع احتمالی بر پایه مشاهدات و دانش پیشین، از پارامترهای به دست می‌دهد. در فصل آینده بر قابلیت‌ها و نتایج این الگوریتم تأکید خواهد شد، و مقدماتً قابل استفاده بودن این الگوریتم به ترتیب چیده خواهد شد. اما ایدهٔ دوم، بی‌ارتباط با الگوریتم *IVB* نیست، و در واقع توزیعی است که در نقش مادهٔ اولیه به این الگوریتم داده می‌شود و مدل مشاهدهٔ ما نیز بر پایهٔ آن شکل گرفته است. ارتباط این دو ایده، پایهٔ تئوریک این پروژه را فراهم می‌کند. به جز آن، هضم مدل فیزیکی مسأله به تدریج در فصل‌های ۲ و ۳ تکمیل می‌شود، و در چند مورد هم از مشاهدات تجربی و شهودی برای ساده‌سازی آن کمک گرفته شده است.

مرور پیشینهٔ تاریخی بحث را از استخراج ویژگی‌ها آغاز می‌کنیم. پس از آنکه دنیس گابور تئوری خود را در سال ۱۹۴۶ در کتاب تئوری مخابرات خویش و در ادامهٔ کارهای نایکوئیست و شانون، در مورد تجزیهٔ هر سیگنال بر اساس واحدهای پایه که در زمان و فرکانس محدود بودند از دیدگاه اصل عدم قطعیت ارائه داد، گرانلوند (*Granlund*) در ۱۹۷۸، فرم کلی یک عملگر پیش‌پردازش تصویری را ارائه داد که بسیار مشابه با توابع سازندهٔ گابور در دو بعد بود، و تنها برحسب نیازهای آن زمان در زمینهٔ

پردازش تصویر و بدون ارتباط با کار گابور به‌وجود آمده بود. وی، به برخی جزئیات همچون محاسبهٔ فواصل توابع پایه نیز پرداخت. علی‌رغم ابداع وی، بیش‌تر تحقیقات انجام شده در این زمینه پس از داومن (*Daugman*) شروع به رشد کردند. وی اولین کسی بود که اصل عدم قطعیت را در دو بعد استخراج نمود و شباهت عجیب میان ساختار قابل اجرا توسط توابع گابور و سیستم بینایی در پستانداران را نمایان ساخت. پس از آن، استفاده از فیلترهای گابور در شناسایی و تفکیک بافت‌ها توسط محققان بسیاری آزموده شد، چرا که بافت‌ها ویژگی‌های تناوبی آماری و ساختاری از خود نشان می‌دهند که توسط توابع گابور قابل توصیف است. تَن (*Tan*) نیز، در راستای مقالهٔ داومن [۳] و ترنر (*Turner*) [۴]، روشی به منظور طبقه‌بندی و نیز تفکیک بافت‌ها ارائه کرد که از ۱۶ فیلتر گابور در چهار جهت و چهار دامنه برای استخراج ویژگی‌ها بهره می‌برد. اما اولین کاری که به طور جامع ایدهٔ بانک فیلترهای گابور در آن به صورت تحلیلی ارائه شده، توسط بویک (*Bovic*) و دیگران [۵] انجام گرفته است. مسألهٔ طراحی و انتخاب پارامترهای فیلتر گابور به نحو موفقیت‌آمیزی توسط جین (*Jain*) و فرخ‌نیا (*Farrokhnia*) در [۶] بررسی شده است. در [۷] نیز، یک روش دیگر بر مبنای استفادهٔ بهینه کانال‌های گابور از فضای فرکانسی ارائه شده است. مسألهٔ شناسایی فونت‌ها اولین بار توسط ژو (*Zhu*) و دیگران [۸] از دیدگاه کانال‌های گابور بررسی شده است و در آن از خروجی فیلتر ۱۶ کانالهٔ گابور به منظور استخراج ویژگی‌های میانگین و واریانس از متن‌های نوشته شده با یک نوع فونت خاص استفاده شده است، که منشاء ایدهٔ این پروژه در استفاده از گابور برای بیان ویژگی‌های مرتبط با نوع قلم نیز بوده است. در تحقیقات مشابهی، به بهبود این ایدهٔ پایه از روش‌هایی چون به کارگیری توابع تعمیم‌یافته (*Grating Cell*) به جای پوش سادهٔ گوسی[۹] و یا استخراج ویژگی‌های پیچیده‌تری مانند گشتاورهای از مرتبهٔ بالاتر [۱۰]،[۱۱] دست زده شده است. شناسایی فونت در [۱۲] بدون استفاده از فیلترهای گابور و تنها با به کارگیری گشتاورهای مرتبهٔ سوم و چهارم روی تصاویر یکنواخت شده‌ای از متن‌های شامل یک نوع فونت، تحقق یافته است. هم‌چنین، در [۱۳]، از هیستوگرام‌های افقی و عمودی در بدست آوردن توزیع یک بعدی و بررسی چولگی و پیچش کلماتِ انگلیسیِ تایپی به منظور طبقه‌بندی برحسب نوع قلم، استفاده شده است. در روش مشابهی در [۱۴]، جهت استخراج ویژگی‌ها از حروف عربی و با هدف تفکیک کلمات به حروف، از شاخصه‌های تخت‌بودن و چولگی هیستوگرام‌های مذکور، به کار گرفته شده است.