

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه صنعتی اصفهان
دانشکده برق و کامپیوتر

توسعه‌ی ماشین بلتزن محدود برای مدل‌سازی مشترک موضوع و احساس در داده‌های متنی

پایان‌نامه کارشناسی ارشد مهندسی کامپیوتر - هوش مصنوعی

مسعود فاطمی

استاد راهنما

دکتر مهران صفایانی

استاد مشاور

دکتر عبدالرضا میرزایی

۱۳۹۵



دانشگاه صنعتی اصفهان
دانشکده برق و کامپیوتر

پایان نامه کارشناسی ارشد رشته مهندسی کامپیوتر – هوش مصنوعی و رباتیک
مسعود فاطمی
تحت عنوان
توسعه‌ی ماشین بلتزن محدود برای مدل سازی مشترک موضوع و احساس در داده‌های متنی

در تاریخ ۹۶/۳/۲۷ توسط کمیته‌ی تخصصی زیر مورد بررسی و تصویب نهایی قرار گرفت:

۱- استاد راهنمای پایان نامه دکتر مهران صفایانی

۲- استاد مشاور پایان نامه دکتر عبدالرضا میرزایی

۳- استاد داور دکتر محمدرضا احمدزاده

۴- استاد داور (اختیاری) دکتر مریم ذکری

سرپرست تحصیلات تکمیلی دانشکده دکتر محمد رضا تابان

تشکر و قدردانی

پروردگار منّان را سپاسگزارم

کلیه حقوق مادی مترتب بر نتایج مطالعات،
ابتکارات و نوآوری‌های ناشی از تحقیق
موضوع این پایان‌نامه متعلق به دانشگاه
صنعتی اصفهان است.

تقدیم به
پدر و مادر عزیزم

فهرست مطالب

عنوان	صفحه
چکیده	یک
فصل اول : مقدمه	۲
۱-۱ اهداف پایان نامه	۳
۲-۱ نوآوری های پایان نامه	۴
۳-۱ مروری بر فصل های پایان نامه	۴
فصل دوم : مفاهیم پایه	۶
۱-۲ مقدمه	۶
۲-۲ متغیر مشاهده شده و متغیر پنهان	۶
۳-۲ مدل احتمالاتی	۷
۲-۳-۱ مدل احتمالاتی افتراقی	۷
۲-۳-۲ مدل احتمالاتی مولد	۸
۴-۲ مدل گرافی	۹
۵-۲ مدل مخلوط	۹
۶-۲ تابع لجستیک	۱۰
۷-۲ Softmax تابع	۱۰
۸-۲ تابع گاما	۱۱
۹-۲ زنجیره ی مارکوف مونت کارلو	۱۱
۱۰-۲ الگوریتم یادگیری واگرایی مقابله	۱۱
۱۰-۲ الگوریتم واگرایی مقابله چیست و چرا به آن احتیاج داریم	۱۲
۱۰-۲ الگوریتم واگرایی مقابله چگونه کار می کند	۱۵
۱۱-۲ کیسه ی کلمات	۱۶
۱۲-۲ موضوع	۱۷
۱۳-۲ مدل موضوعی	۱۷
۱۴-۲ آنالیز احساس	۱۹

۲۱	فصل سوم: بررسی پژوهش‌های پیشین
۲۱	۱-۳ مقدمه
۲۳	۲-۳ مدل‌های تخمین توزیع احتمالی
۲۴	۱-۲-۳ مدل ماشین بلترن محدود
۲۵	۲-۲-۳ مدل شبکه‌ی عصبی خود رگرسیو تخمین‌زننده‌ی توزیع
۲۶	۳-۳ روش‌های مدل‌سازی موضوعی
۲۶	۱-۳-۳ مدل تکرار عبارت-معکوس تکرار سند
۲۷	۲-۳-۳ مدل فهرست‌سازی معنایی نهفته
۲۸	۳-۳-۳ مدل فهرست‌سازی معنایی نهفته‌ی احتمالاتی
۲۹	۴-۳-۳ مدل تخصیص دیریکله‌ی پنهان
۳۱	۵-۳-۳ مدل Softmax تکرار شده
۳۲	۶-۳-۳ مدل شبکه‌ی عصبی خود رگرسیو تخمین‌زننده‌ی توزیع سندی
۳۳	۴-۳ مدل‌های مشترک موضوع و احساس
۳۴	۱-۴-۳ مدل یکی‌سازی احساس-موضوع
۳۵	۲-۴-۳ مدل نظارت‌شده‌ی ضعیف تشخیص مشترک احساس-موضوع
۳۷	۳-۴-۳ مدل نظارت‌شده‌ی مشترک موضوع و احساس
۳۷	۵-۳ مدل‌های چند حالت
۳۸	۱-۵-۳ مدل نظارت‌شده‌ی شبکه‌ی عصبی خود رگرسیو تخمین‌زننده‌ی توزیع سندی
۳۸	۶-۳ نتیجه‌گیری
۴۰	فصل چهارم: مدل پیشنهادی
۴۰	۱-۴ مقدمه
۴۰	۲-۴ مدل پایه
۴۴	۳-۴ ماشین بلترن محدود با واحدهای قابل مشاهده صحیح
۴۷	۴-۴ مدل مولد براساس تعداد کلمات
۴۹	۵-۴ مدل پیشنهادی مولد احتمالی احساس/موضوع
۵۲	۶-۴ نتیجه‌گیری
۵۴	فصل پنجم: شبیه‌سازی و ارزیابی مدل پیشنهادی
۵۴	۱-۵ مقدمه
۵۵	۲-۵ پیش‌پردازش‌های متنی در پردازش زبان طبیعی
۵۵	۱-۲-۵ علامت‌گذاری و حذف کاراکترهای بی‌معنی
۵۵	۲-۲-۵ ریشه‌یابی لغوی
۵۶	۳-۲-۵ ریشه‌یابی نحوی
۵۶	۴-۲-۵ حذف کلمات توقف
۵۶	۳-۵ پایگاه داده‌ی بازبینی فیلم

۵۷	۴-۵ پایگاه داده‌ی ۲۰ گروه خبری
۵۷	۵-۵ پایگاه داده‌ی احساس چند دامنه
۵۷	۶-۵ آماده‌سازی پایگاه داده
۵۹	۷-۵ لغت‌نامه‌ی احساس
۶۰	۸-۵ جزئیات آموزش مدل پیشنهادی
۶۱	۹-۵ مدل‌سازی اسناد و ارزیابی به عنوان یک مدل مولد
۶۳	۱۰-۵ مجسم‌سازی موضوع‌ها و ارزیابی دقت در محاسبه‌ی آن‌ها
۶۵	۱۱-۵ طبقه‌بندی احساسی اسناد
۶۸	۱۲-۵ بازیابی اطلاعات
۷۱	۱۳-۵ نتیجه‌گیری

فصل ششم: نتیجه‌گیری و پیشنهادات

۷۲	۱-۶ نتیجه‌گیری
۷۴	۲-۶ پیشنهادها

مراجع

۷۷	چکیده انگلیسی
----	---------------

فهرست تصاویر

۹	۱-۲ نمونه‌ای از یک مدل گرافی [۱]
۱۰	۲-۲ تابع لجستیک در بازه‌ی اعداد حقیقی از منفی بی نهایت تا مثبت بی نهایت
۱۹	۳-۲ نمونه‌ای از یک مدل‌سازی موضوعی
۲۵	۱-۳ ماشین بلترمن محدود
۲۶	۲-۳ مدل‌های FVBN و NADE [۲]
۲۷	۳-۳ جدول‌های تعداد تکرار کلمات برای یک مجموعه‌ی اسناد ۲ سندی
۲۹	۴-۳ مدل فهرست‌سازی معنایی نهفته‌ی احتمالاتی [۱]
۳۰	۵-۳ مدل تخصیص دیریکله‌ی پنهان [۱]
۳۱	۶-۳ نمونه خروجی مدل LDA برای چهار موضوع مختلف [۱]
۳۲	۷-۳ مدل Softmax تکرار شده [۳]
۳۳	۸-۳ مدل شبکه‌ی عصبی خود رگرسیو تخمین‌زننده‌ی توزیع سندی [۴]
۳۵	۹-۳ مدل یکی‌سازی احساس-موضوع [۵]
۳۶	۱۰-۳ مدل نظارت‌شده‌ی ضعیف تشخیص مشترک احساس/موضوع [۶]
۳۶	۱۱-۳ خروجی مدل JST برای دو احساس مثبت و منفی و دو موضوع برای هر احساس و پانزده کلمه در هر موضوع [۶]
۳۷	۱۲-۳ مدل نظارت‌شده‌ی مشترک موضوع و احساس [۷]
۳۸	۱۳-۳ مدل نظارت‌شده‌ی شبکه‌ی عصبی خود رگرسیو تخمین‌زننده‌ی توزیع سندی [۸]
۴۲	۱-۴ ماشین بلترمن محدود
۴۵	۲-۴ ماشین بلترمن محدود با واحدهای قابل مشاهده صحیح
۵۰	۳-۴ مدل پیشنهادی مولد احتمالی احساس/موضوع با Softmax تکرار شده
۵۰	۴-۴ یک سند ورودی با ۵ کلمه در مدل بخش ۳-۴
۵۹	۱-۵ نمونه‌ای از یک سند منفی در فایل lib-svm
۵۹	۲-۵ نمونه‌ای از لغت‌نامه احساس MPQA شامل ۶ کلمه
۶۲	۳-۵ ارزیابی تغییرات سرگشتگی در فرآیند آموزش بر روی پایگاه داده‌ی MR برای مدل پیشنهادی و مدل RS
۶۵	۴-۵ ارزیابی دقت در تخصیص احساس به موضوع‌ها به ۲۰۰ و ۱۰۰۰ مرحله آموزش
۶۷	۵-۵ طبقه‌بندی احساس در پایگاه داده‌ی MR با استفاده از مدل پیشنهادی و مدل پایه برای موضوع‌های مختلف

- ۵-۶ طبقه‌بندی احساس در پایگاه داده‌ی MR با استفاده از مدل‌های شبکه عصبی با مقدار دهی اولیه برای وزن‌ها و بایاس‌ها، شبکه عصبی و SVM ۶۸
- ۵-۷ بازیابی اطلاعات با استفاده از پایگاه داده‌ی ۲۰ News Groups و MRMDS برای رویکرد پیشنهادی و مدل RS ۷۰

فهرست جداول

۱۷	۱-۲ نمونه‌ی چهار موضوع مختلف
۲۲	۱-۳ دسته‌بندی مدل‌های پیشین از نظر ساختار
۲۳	۲-۳ دسته‌بندی مدل‌های پیشین از نظر نحوه‌ی عملکرد
۲۳	۳-۳ دسته‌بندی مدل‌های پیشین از نظر داده‌ی ورودی
۳۴	۴-۳ نمونه‌ای از موضوع‌های یادگرفته شده توسط مدل DocNADE [۴]
۵۸	۱-۵ اطلاعات آماری پایگاه داده‌ی Movie Review
۶۳	۲-۵ تخمین سرگشتگی برای پایگاه داده‌ی Movie Review با استفاده از مدل پیشنهادی
۶۴	۳-۵ فراوانی‌های بدست آمده از مقایسه‌ی کلمات مشترک بین لغت‌نامه‌ی احساسی MPQA با سه لغت‌نامه واژگان

چکیده

امروزه با گسترش اینترنت و وب، انواع مختلف رسانه‌های اجتماعی نظیر وبلاگ‌ها و شبکه‌های اجتماعی به یک منبع بسیار عظیم از داده‌های متنی تبدیل شده‌اند. با پردازش این داده‌ها می‌توان اطلاعات سودمند و مفیدی در مورد مباحث مختلف، نظر افراد و احساس کلی جامعه بدست آورد. از این جهت داشتن مدل‌هایی که کاملاً خودکار به تشخیص اطلاعات مفهومی و احساس در اسناد متنی بپردازند بسیار مفید است. روش‌های مدل‌سازی موضوع و استخراج اطلاعات مفهومی و همچنین تشخیص احساس از مهم‌ترین مباحث مطرح شده در زمینه‌ی پردازش زبان طبیعی، و کاوش داده‌های متنی هستند. بیشتر مدل‌هایی که در این زمینه وجود دارند بر پایه‌ی روش‌های آماری و شبکه‌های بیزی هستند به طوری که در زمینه‌ی مدل‌سازی موضوع-احساس با استفاده از شبکه‌های عصبی تا به امروز هیچ رویکردی وجود ندارد. همچنین بیشتر رویکردهای موجود دارای محدودیت‌هایی مانند پیچیدگی محاسباتی بالا هستند. در این پایان‌نامه یک ساختار جدید برای مدل‌سازی مشترک احساس-موضوع در داده‌های متنی بر پایه‌ی شبکه‌ی عصبی ماشین بلتزن محدود پیشنهاد می‌گردد. با تغییر ساختار ماشین بلتزن محدود و اضافه کردن یک لایه به آن که متناظر با احساس اسناد متنی است یک ساختار مولد احتمالی برای مدل‌سازی مشترک احساس و موضوع بر پایه‌ی شبکه‌ی عصبی پیشنهاد می‌دهیم. رویکرد پیشنهاد شده یک رویکرد نظارت شده است که برای آموزش آن از الگوریتم واگرایی مقابله استفاده می‌کنیم. لایه‌ی جدید اضافه شده در مدل پیشنهادی لایه‌ای با ماهیت توزیع احتمالی چند جمله‌ای است که از آن می‌توان در فرآیند عمل طبقه‌بندی اسناد متنی از نظر احساسی یا دیگر کاربردهای نظارت شده استفاده کرد. مدل پیشنهادی پس از پیاده‌سازی در فرآیندهای: مدل‌سازی به عنوان یک مدل مولد، طبقه‌بندی احساس و بازیابی اطلاعات با مدل‌های موجود مقایسه گردید و نتایج بدست آمده به طور کامل گزارش شده است. مشاهده گردید در فرآیند طبقه‌بندی احساس مدل پیشنهادی به طور میانگین ۱۱ درصد دقت بهتری نسبت به مدل پایه دارد. همچنین در فرآیند بازیابی اطلاعات بر روی پایگاه داده‌ی ۲۰ گروه خبری، رویکرد پیشنهادی با در نظر گرفتن احساس به طور متوسط ۹ درصد عملکرد بهتری نسبت به حالتی که در آن احساس در نظر گرفته نمی‌شود را به همراه دارد.

کلمات کلیدی: مدل‌سازی موضوع، آنالیز احساس، شبکه‌های عصبی، ماشین بلتزن محدود، مدل احتمالاتی، الگوریتم واگرایی مقابله

فصل اول

مقدمه

”آیا ماشین‌ها توانایی فکر کردن و یادگیری را دارند؟“ اولین بار این سوال در سال ۱۹۵۰ میلادی تحت عنوان یک مقاله با همین عنوان توسط ”آلن تورینگ“ که امروزه از او به عنوان پدر علم هوش مصنوعی^۲ یاد می‌گردد در یک مجله‌ی فلسفی مطرح گردید. با گذر زمان و پیدایش سیستم‌های کامپیوتری پیچیده و مطرح شدن مباحث مربوط به هوش مصنوعی به شکل امروزی، این سوال همواره به عنوان بزرگترین و چالش برانگیزترین سوال برای محققین و کارشناسان این حوزه مطرح بوده است. امروزه در تمام مباحث مربوط به هوش مصنوعی ما به دنبال روش‌ها، الگوریتم‌ها و ساختارهایی هستیم که بتوانند هرچه بهتر، به صورت خودکار و با دقت بالا یک رفتار انسانی و یا فرا انسانی را با بیشترین سرعت ممکن انجام دهند. اعمالی مانند دسته‌بندی^۳، استخراج اطلاعات مفهومی^۴، تحلیل^۵ و برچسب گذاری^۶ داده‌ها و از جمله فعالیت‌هایی هستند که امروزه ما انجام بسیاری از آن‌ها را به ماشین‌ها واگذار می‌کنیم.

در بین انواع مختلف داده شاید به جرات بتوان بیان کرد که داده‌های متنی^۷ همواره دارای سهم عظیمی از

^۲Artificial Intelligence

^۳Classification

^۴Subjective Information Extraction

^۵Analysis

^۶Labeling

^۷Text Data

نظر حجم و مقدار هستند. به خصوص با گسترش اینترنت و وب در دهه‌ی اخیر با سرعتی بسیار زیاد، انواع مختلف رسانه‌های اجتماعی نظیر وبلاگ‌ها، شبکه‌های اجتماعی و گروه‌های بحث در اینترنت به یک منبع بسیار عظیم و قوی از انواع مختلف داده و اطلاعات به ویژه داده‌های متنی تبدیل شده اند که با پردازش این داده‌ها می‌توان اطلاعات سودمند و مفیدی در مورد مباحث مختلف، نقطه نظر افراد و احساس کلی جامعه بدست آورد [۶]، چرا که درک کردن و فهمیدن اینکه دیگر افراد چگونه فکر می‌کنند همواره یک هدف و بخشی بسیار مهم در بحث جمع‌آوری اطلاعات است [۹].

در مباحث مربوط به حوزه هوش مصنوعی، فعالیت‌های انجام گرفته در زمینه کاوش داده‌ها^۱ به خصوص کاوش داده‌های متنی^۲ و همچنین پردازش زبان طبیعی^۳، بیشتر از هر زمینه‌ی دیگری به تلاش برای درک و فهم این حجم عظیم از داده‌های متنی مربوط می‌شوند. حجمی عظیم از داده‌های متنی که بدون هیچ ساختار و قاعده و قانونی هستند و روز به روز مقدار آن‌ها با سرعت بسیاری چشمگیری در حال افزایش است. در این میان وجود الگوریتم‌ها و روش‌هایی که بتوانند به صورت خودکار با این حجم بسیار زیاد از داده‌های بدون ساختار ارتباط برقرار کرده و اطلاعات مفید و سودمند را از آن برای ما استخراج کنند بیش از پیش احساس می‌گردد.

روش‌های مدل‌سازی موضوع^۴ و استخراج اطلاعات مفهومی از داده‌های ورودی به خصوص داده‌های متنی و همچنین تشخیص احساس^۵، همواره از مهمترین مباحث مطرح شده در زمینه‌ی پردازش زبان طبیعی و کاوش داده‌های متنی بوده است. این روش‌ها که اکثراً در دسته‌ی روش‌های بدون نظارت^۶ قرار می‌گیرند با اجرا بر روی یک پایگاه داده از داده‌های متنی توانایی تشخیص و مدل‌سازی موضوعات و مفاهیم همراه با هر سند متنی را دارا هستند. تشخیص احساس برای هر سند و هر موضوع در بحث بازیابی اطلاعات^۷ (IR) نیز می‌تواند به اندازه تشخیص اطلاعات موجود در هر متن حائز اهمیت باشد. از این جهت داشتن مدل‌هایی که به صورت اتوماتیک و کاملاً خودکار به مدل‌سازی موضوع و تشخیص اطلاعات مفهومی و احساس در اسناد بپردازند می‌تواند بسیار مفید باشد.

۱-۱ اهداف پایان‌نامه

کاری که ما در این پایان‌نامه قصد انجام آن را داریم نیز در همین راستا است و هدف ارائه‌ی روشی بر پایه‌ی شبکه‌های عصبی مصنوعی^۸ برای استخراج موضوع‌های مختلف و احساسات همراه با آن‌ها در یک مجموعه

^۱Data Mining

^۲Text Mining

^۳Natural Language Processing

^۴Topic Model

^۵Sentimen Detection

^۶Unsupervised

^۷Information Retrieval

^۸Artificial Neural Networks

از داده‌های متنی است. بیشتر کارهایی که در این زمینه وجود دارد بر پایه‌ی مدل‌های آماری و شبکه‌های بیزی^۱ هستند که دچار پیچیدگی محاسباتی هستند. در بحث شبکه‌های عصبی مصنوعی بر خلاف مدل‌های آماری، روش‌های زیادی وجود ندارند که برای ما مدل کردن موضوع و احساس را انجام دهند. لذا مدلی که در این پایان‌نامه پیشنهاد شده دارای رویکردی جدید برای یک مسأله‌ی جدید است که پس از پیاده‌سازی و آزمایش بر روی پایگاه داده‌های مختلف با مدل‌های موجود مقایسه می‌شود.

تمرکز ما در انجام این پایان‌نامه و روش پیشنهادی پردازش بر روی داده‌های متنی است. در تقابل با داده‌های متنی هدف پیدا کردن توزیع موضوع‌های مختلف موجود در مجموعه اسناد پایگاه داده و همچنین توزیع کلمات و احساس همراه با هر موضوع با استفاده شبکه‌های عصبی مصنوعی است. موضوع کلی و فرآیند مورد نظر در داده‌های متنی تحت عنوان مدل کردن موضوع شناخته می‌شود که در مباحث مربوط به هوش مصنوعی در دسته‌ی کارهای مربوط به یادگیری ماشین، پردازش زبان طبیعی، شبکه‌های عصبی مصنوعی و کاوش احساسات قرار می‌گیرد.

۲-۱ نوآوری‌های پایان‌نامه

در بحث مدل‌سازی موضوع با استفاده از شبکه‌های عصبی در سال‌های اخیر تعداد اندکی روش ارائه شده است. اما در زمینه‌ی مدل‌سازی مشترک احساس و موضوع با استفاده از شبکه‌های عصبی تا کنون هیچ مدلی مطرح نشده و مورد آزمایش قرار نگرفته است. نتایج بهتر مدل‌های شبکه عصبی در بحث مدل‌سازی موضوع در مقایسه با روش‌های پیشین که از ساختارهای گرافی و مدل‌های بیزی استفاده می‌کردند، همچنین عدم وجود روشی برای تشخیص همزمان احساس و موضوع در داده‌های متنی با استفاده از شبکه‌های عصبی منجر به رویکرد پیشنهادی در این پژوهش برای مدل‌سازی مشترک احساس و موضوع در داده‌های متنی بر پایه‌ی شبکه‌های عصبی گردید.

۳-۱ مروری بر فصل‌های پایان‌نامه

در این تحقیق ابتدا در فصل دوم به معرفی مفاهیم پایه می‌پردازیم. منظور از مفاهیم پایه تعاریف و مفاهیم اولیه و مورد نیاز در حوزه‌ی مدل‌های احتمالاتی، شبکه‌های عصبی و مدل‌سازی احساس و موضوع هستند. در این فصل دو دسته‌ی مهم از مدل‌های احتمالاتی یعنی مدل‌های مولد و افتراقی را تعریف می‌کنیم و بیان می‌کنیم که مدل پیشنهادی در این پایان‌نامه در کدام دسته مدل‌ها قرار می‌گیرد. همچنین یک الگوریتم آموزش برای دسته‌ی خاصی از شبکه‌های عصبی و همچنین توابع و مفاهیم پایه در بحث مدل‌سازی موضوع را به صورت کامل تعریف

^۱ Bayesian Networks

می‌کنیم.

در ادامه و در فصل سوم به مرور کارهای پیشین در زمینه‌ی تخمین توزیع‌های احتمالی در داده‌های ورودی، مدل‌سازی احساس و مدل‌سازی احساس-موضوع در داده‌های متنی می‌پردازیم. ضمن تعریف مهم‌ترین مدل‌های موجود در این زمینه و بیان نقاط ضعف و قدرت آن‌ها توضیح می‌دهیم که ایده‌ی این پژوهش از کجا و به چه دلیل شکل گرفته و نسبت به مدل‌های پیشین از چه مزیت‌هایی برخوردار است.

در فصل چهارم کلیات نظری و تئوری مدل پیشنهادی در این پژوهش به صورت دقیق بیان می‌شوند. با بیان کاستی‌های رویکردهای موجود در حوزه‌ی مدل‌سازی احساس و موضوع با استفاده از شبکه‌های عصبی یک روش جدید برای این منظور پیشنهاد می‌گردد. در این فصل با معرفی یک مدل معروف به عنوان پایه و زیرساخت و چند روش گسترش یافته از آن ساختار مدل جدید تعریف و قسمت‌های مختلف آن شرح داده می‌شوند و روابط مورد نیاز برای هر بخش به صورت دقیق تعریف می‌شوند.

در فصل پنجم مراحل شبیه‌سازی مدل پیشنهادی و نتایج حاصل از آن به طور مفصل توضیح داده خواهد شد. در ادامه و در فصل آخر، نتیجه‌گیری حاصل از انجام این پژوهش شرح داده خواهد شد. همچنین راهکارهایی برای بهبود و توسعه مدل پیشنهادی ارائه خواهد شد.

فصل دوم

مفاهیم پایه

۱-۲ مقدمه

در این قسمت به بررسی و توضیح مفاهیم پایه مورد نیاز در بحث مدل سازی موضوع و آنالیز احساس^۲ می پردازیم. در ابتدا بر روی مفاهیم ریاضی و مدل های احتمالاتی^۳، انواع و تفاوت های آنها تمرکز می کنیم، سپس در مورد توابع ریاضی مورد نیاز در بخش های بعدی توضیح داده می شود و در ادامه با مفاهیم موجود در حوزه ی مدل سازی موضوعی و کاوش عقاید^۴ آشنا می شویم و در مورد هر کدام توضیحات مورد نیاز را ارائه می دهیم.

۲-۲ متغیر مشاهده شده و متغیر پنهان

در علم آمار^۵ و احتمال متغیر مشاهده شده^۶ به متغیری گفته می شود که مقدار آن واقعا مشاهده شده باشد و یا واقعا اتفاق افتاده باشد. در مقابل آن متغیر مخفی یا پنهان^۷ وجود دارد که به متغیری گفت می شود که از دیگر متغیرهای مشاهده شده استنتاج^۸ می شود.

^۲Sentiment Analysis

^۳Probabilistic Model

^۴Opinion Mining

^۵Statistic

^۶Observed Variable

^۷Latent Variable

^۸Inference

۲-۳ مدل احتمالاتی

به زبان ساده، یک مدل احتمالی که به آن مدل آماری^۱ نیز می‌گویند، دسته‌ای از مدل‌های ریاضی است که با مجموعه‌ای از فرضیات همراه است که بر روی تولید داده‌های نمونه^۲ و یا داده‌های مشابه از یک جمعیت بزرگتر یا از یک سیستم تمرکز می‌کنند. فرضیاتی که در اینجا در مورد آن صحبت می‌کنیم و همراه با یک مدل احتمالی هستند در واقع توزیع‌های احتمالاتی^۳ هستند که احتمال رخداد یا تولید نمونه‌های مختلف از جمعیت بزرگتر را برای ما مشخص می‌کنند. به بیان دیگر مدل احتمالی به مدلی گفته می‌شود که داده‌هایی که می‌توانند از یک سیستم مشاهده شوند را برای ما توصیف می‌کند.

به بیان ریاضی، یک مدل احتمالی به صورت یک جفت به شکل (S, P) در نظر گرفته می‌شود که در آن S مجموعه‌ی تمام مشاهدات ممکن یا به عبارت دیگر همان فضای نمونه^۴ و P مجموعه‌ی توزیع‌های احتمالاتی بر روی S می‌باشد. در این تعریف فرض بر این است که یک توزیع احتمالاتی درست منجر به تولید داده‌های قابل مشاهده گردیده است. ما P را به عنوان مجموعه‌ای از توزیع‌های احتمالاتی انتخاب می‌کنیم که به اندازه‌ی کافی این توزیع احتمالی درست را با دقت مناسب تقریب بزنیم.

مدل‌های احتمالی را از نظر توزیع احتمالی که سعی در تقریب زدن آن به شکلی مناسب و با دقت کافی دارند را می‌توان در دو دسته کلی مدل‌های مولد^۵ و مدل‌های افتراقی^۶ دسته‌بندی کرد، که در ادامه به توضیح هر کدام از این دو دسته می‌پردازیم و آن‌ها را با یکدیگر مقایسه می‌کنیم.

۲-۳-۱ مدل احتمالاتی افتراقی

مدل‌های افتراقی که همچنین از آن‌ها به عنوان مدل‌های شرطی^۷ نیز یاد می‌شود یک کلاس از مدل‌های احتمالی است که بیشتر در مباحث مربوط به یادگیری ماشین^۸ برای مدل کردن وابستگی یک متغیر مشاهده نشده مانند y به یک متغیر مشاهده شده مانند x مورد استفاده قرار می‌گیرند. در یک مدل افتراقی $p(y|x)$ که یک توزیع احتمالی شرطی است مدل می‌شود و می‌توان از آن برای پیش‌بینی y با توجه به x استفاده کرد.

برای کاربردهایی مانند طبقه‌بندی و رگرسیون^۹ که در آن‌ها از احتمالات شرطی استفاده می‌کنیم مدل‌های افتراقی نسبت به مدل‌های دیگر نتایج بهتری را از خود نشان می‌دهند. علاوه بر این مدل‌های افتراقی به صورت

¹Statistical Model

²Sample Data

³Probabilistic Distribution

⁴Sample Space

⁵Generative Models

⁶Discriminative Models

⁷Conditional Models

⁸Machine Learning

⁹Regression

ذاتی در دسته مدل‌های نظارت‌شده^۱ قرار می‌گیرند و به آسانی قابل بسط دادن به حالت بدون نظارت ناست. مدل‌هایی مانند: ماشین‌های بردار پشتیبان^۲، شبکه‌های عصبی^۳، رگرسیون خطی^۴ و غیره که در یادگیری ماشین از آن‌ها استفاده می‌کنیم در دسته‌ی مدل‌های افتراقی قرار می‌گیرند.

۲-۳-۲ مدل احتمالاتی مولد

مدل‌های احتمالاتی مولد در مقابل مدل‌های احتمالاتی افتراقی قرار دارند و در مسائل یادگیری پیچیده از انعطاف‌پذیری بیشتری نسبت به مدل افتراقی برخوردارند. به طور کلی در مباحث آماری و احتمالاتی، یک مدل مولد به مدلی گفت می‌شود که با در نظر گرفتن تعدادی متغیر پنهان داده شده، داده‌های قابل مشاهده را تولید می‌کند. اساس کار مدل‌های مولد توزیع احتمالاتی مشترک^۵ است. این مدل‌ها نیز همانند دسته‌ی پیشین در مباحث مربوط به یادگیری ماشین به دو شکل کلی مورد استفاده قرار می‌گیرند. (۱) برای مدل کردن داده‌ها به صورت مستقیم یا به عبارت دیگر مدل کردن داده‌های قابل مشاهده که از یک تابع چگالی احتمال^۶ بدست می‌آیند. (۲) به عنوان یک مرحله‌ی میانی برای تشکیل یک تابع چگالی احتمال شرطی^۷ که با استفاده از قانون بیز^۸ ساخته می‌شوند مورد استفاده قرار می‌گیرند.

یک مدل مولد، یک مدل احتمالاتی کامل بر اساس تمام متغیرها است. در حالی که یک مدل افتراقی، تنها یک مدل مشروط برای متغیر هدف^۹ بر اساس متغیرهای قابل مشاهده را مهیا می‌کند، لذا از این نظر مدل‌های افتراقی و مولد کاملاً در مقابل یکدیگر قرار دارند. در نتیجه مدل‌های مولد برای مواردی نظیر شبیه‌سازی یا تولید هر یک از متغیرها در یک مدل می‌توانند مورد استفاده قرار بگیرند، در حالی که مدل‌های افتراقی تنها اجازه‌ی نمونه‌برداری^{۱۰} از متغیر هدف مشروط به مقادیر قابل مشاهده را می‌دهند. خصوصیات بیان شده برای مدل‌های مولد باعث می‌شود که این مدل به نتایج بهتری در مسائل بدون نظارت دست یابند. مدل‌هایی مانند: مدل مخلوط گوسی^{۱۱}، مدل مخفی مارکوف^{۱۲}، تخصیص دیریکله‌ی پنهان^{۱۳} که در فصل بعدی به صورت کامل توضیح داده می‌شود و ماشین بلتزمن محدود^{۱۴} که در فصل مدل پیشنهادی از آن استفاده می‌کنیم در دسته‌ی مدل‌های مولد قرار می‌گیرند.

¹Supervised

²Support Vector Machine

³Neural Networks

⁴Linear Regression

⁵Joint Probability Distribution

⁶Probability Density Function

⁷Conditional Probability Density Function

⁸Bayes Rule

⁹Target Variable

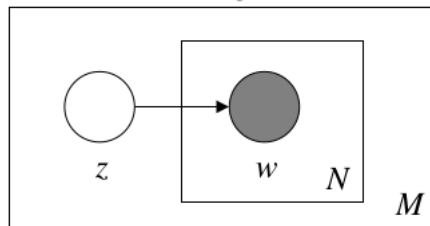
¹⁰Sampling

¹¹Gaussian Mixture Model

¹²Hidden Markov Model

¹³Latent Dirichlet Allocation

¹⁴Restricted Boltzman Machine



شکل ۲-۱: نمونه‌ای از یک مدل گرافی [۱]

۴-۲ مدل گرافی

مدل‌های گرافی^۱ یا مدل‌های گرافی احتمالاتی، دسته‌ای از مدل‌های احتمالاتی هستند که در حالت کلی برای نمایش وابستگی شرطی بین متغیرهای تصادفی در آن‌ها، از یک نمودار گراف مانند استفاده می‌کنند. کاربرد اصلی این مدل‌ها در بحث تئوری آمار و احتمال (به‌خصوص آمار و احتمال بیزی) و یادگیری ماشین است. در شکل ۲-۱ یک نمونه از مدل‌های گرافی را مشاهده می‌کنید. در این شکل و به طور کلی در این مدل‌ها، مستطیل‌ها نشان دهنده تکرار و تعداد هستند. پیکان‌های جهت‌دار وابستگی‌های شرطی را نشان می‌دهند و دایره‌ها نشان دهنده متغیرهای تصادفی هستند. متغیرهای تصادفی می‌توانند سفید (پنهان و مشاهده نشده) و یا تیره (متغیرهای قابل مشاهده) باشند.

برای مثال در شکل ۲-۱ که یک مدل گرافی ساده را نشان می‌دهد، مستطیل بیرونی M بار و مستطیل درونی به ازای هر بار تکرار مستطیل بیرونی N بار تکرار می‌شود. همچنین این مدل از دو متغیر تصادفی تشکیل شده که بین آن‌ها یک وابستگی شرطی نیز وجود دارد و یکی از آن‌ها یک متغیر پنهان (z) و دیگری یک متغیر قابل مشاهده (w) می‌باشد.

۵-۲ مدل مخلوط

در علم آمار یک مدل مخلوط^۲ یک مدل احتمالاتی است که از آن برای نشان دادن وجود چند زیر جامعه^۳ در یک جامعه بزرگتر بدون نیاز به داده‌ی قابل مشاهده برای شناسایی و تشخیص آن زیر جامعه‌ها استفاده می‌شود. به بیان رسمی، یک مدل مخلوط متناظر با یک توزیع مخلوط^۴ است که توزیع‌های احتمالاتی مشاهده شده در یک جامعه کلی یا جمعیت بزرگتر را نشان می‌دهد.

¹ Graphical Model

² Mixture Model

³ Subpopulation

⁴ Mixture Distribution

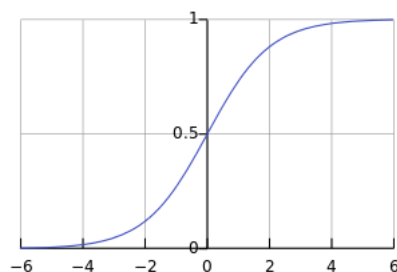
۶-۲ تابع لجستیک

تابع لجستیک^۱ یک منحنی متداول از خانواده‌ی منحنی‌های سیگمویدی^۲ است. معادله‌ی:

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}} \quad (۱-۲)$$

نشان دهنده یک تابع لجستیک است که در آن x_0 نقطه میانی منحنی سیگمویدی روی محور x ها، L مقدار بیشینه منحنی در مثبت بی نهایت و k برابر با شیب منحنی است.

در شکل ۲-۲ برای مقادیر x در بازه‌ی اعداد حقیقی از منفی بی نهایت تا مثبت بی نهایت نمودار این منحنی نشان داده شده است. تابع لجستیک در ادامه برای معرفی تابع softmax مورد استفاده قرار می‌گیرد.



شکل ۲-۲: تابع لجستیک در بازه‌ی اعداد حقیقی از منفی بی نهایت تا مثبت بی نهایت

۷-۲ تابع Softmax

در ریاضیات تابع Softmax که به آن تابع توانی نرمال‌شده^۳ نیز می‌گویند، یک تابع لجستیک تعمیم یافته است که از آن برای نرم کردن یک بردار K بعدی مانند \mathbf{z} متشکل از اعداد حقیقی دلخواه به یک بردار K بعدی مانند $\sigma(\mathbf{z})$ تشکیل شده از اعداد حقیقی بین بازه‌ی $(0, 1)$ به گونه‌ای که مجموع آنها برابر با ۱ باشد استفاده می‌کنند. رابطه‌ی:

$$\sigma(\mathbf{z}) = \frac{e^{z_i}}{\sum_{k=1}^K e^{z_k}} \quad for \quad i = 1, \dots, K \quad (۲-۲)$$

معادله‌ی تابع Softmax را نشان می‌دهد. با توجه به خاصیت ذکر شده برای بردار حاصل از تابع Softmax در تئوری احتمال از این خروجی برای تعریف یک توزیع احتمالی که دارای K خروجی ممکن است استفاده می‌کنند. در بخش‌های بعدی از این تابع برای تعریف توزیع‌های احتمالی برای موضوعات مختلف استفاده می‌کنیم.

¹Logistic Function

²Sigmoid Curves

³Normalized Exponential Function

۸-۲ تابع گاما

تابع گاما^۱ که آن را با Γ نشان می‌دهند از تابع فاکتوریل مشتق شده است و برای یک عدد صحیح مثبت مقدار آن از رابطه:

$$\Gamma(n) = (n-1)! \quad (۳-۲)$$

بدست می‌آید. تابع گاما در واقع برای تمام اعداد نامنفی حتی اعداد مختلط نیز تعریف می‌شود. برای یک عدد مختلط با بخش حقیقی مثبت مقدار این تابع از رابطه

$$\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx \quad (۴-۲)$$

بدست می‌آید.

۹-۲ زنجیره‌ی مارکوف مونت کارلو

روش زنجیره مارکوف مونت کارلو^۲ که به اختصار آن را MCMC می‌گویند، در کلاس روش‌های تولید نمونه از یک توزیع احتمالاتی قرار دارد. به زبان خیلی ساده MCMC با ساختن یک زنجیر مارکوف برگشت‌پذیر^۳ که در حالت تعادل توزیعی مشابه توزیع مورد نظر ما را دارد، از توزیع احتمالاتی نمونه تولید می‌کند. به بیان دیگر حالت زنجیر مارکوف، بعد از طی کردن چند مرحله به عنوان یک نمونه از توزیع احتمالاتی مورد نظر در نظر گرفته می‌شود. کیفیت نمونه‌ی تولید شده توسط روش MCMC با تعداد مراحل که زنجیر مارکوف اجرا می‌گردد رابطه‌ی مستقیم دارد و هرچه تعداد این مراحل بیشتر باشد نمونه‌ی بهتری بدست خواهد داد.

۱۰-۲ الگوریتم یادگیری واگرایی مقابله

الگوریتم یادگیری واگرایی مقابله^۴ اولین بار در سال ۲۰۰۲ توسط آقای هیئتون [۱۰] به عنوان یک الگوریتم یادگیری تقریبی بیشینه‌سازی مورد انتظار^۵ پیشنهاد گردید که در اینجا آن را به طور کامل توضیح می‌دهیم، و در ادامه خواهیم دید که پایه و اساس روش‌های یادگیری مدل‌های پیشین و پیشنهادی در این پایان‌نامه بر اساس همین الگوریتم هستند.

^۱Gamma Function

^۲Markov Chain Mont Carlo

^۳Reversible

^۴Contrastive Divergence

^۵Maximum Likelihood Approximation

۲-۱۰-۱ الگوریتم واگرایی مقابله چیست و چرا به آن احتیاج داریم

فرض کنید که قصد مدل سازی احتمال یک نقطه داده^۱ مانند x را با استفاده از یک تابع به شکل $f(x; \theta)$ که در آن θ یک بردار از پارامترهای مدل است را داریم. می دانیم که احتمال x که آن را به صورت $p(x; \theta)$ نمایش می دهیم به ازای مجموع تمام حالت های x باید برابر با ۱ شود. بنابراین داریم:

$$p(x; \theta) = \frac{1}{Z(\theta)} f(x; \theta) \quad (۵-۲)$$

که در آن $Z(\theta)$ تابع قسمت بندی^۲ است و به فرم

$$Z(\theta) = \int f(x; \theta) dx \quad (۶-۲)$$

تعریف می شود. تابع قسمت بندی برای ما تضمین می کند که مقدار بدست آمده برای عبارت سمت چپ رابطه ی $۵-۲$ یک مقدار صحیح احتمالی (بین ۰ و ۱) است [۱۱].

مجموعه پارامترهای مدل را که در اینجا با θ نشان می دهیم، را با بیشینه کردن احتمال یک مجموعه ی آموزش^۳ از داده ها که آن را به صورت $\mathbf{X} = x_{1,...,K}$ تعریف می کنیم، یاد می گیریم. که در این حالت احتمال مجموعه آموزش از رابطه ی

$$p(\mathbf{X}; \theta) = \prod_{k=1}^K \frac{1}{Z(\theta)} f(x_k; \theta) \quad (۷-۲)$$

بدست می آید. یا با کمینه کردن مقدار منفی لگاریتم $p(\mathbf{X}; \theta)$ که آن را با $E(\mathbf{X}; \theta)$ نشان داده و انرژی مربوط به مدل تعریف می کنیم و از رابطه ی

$$E(\mathbf{X}; \theta) = \log Z(\theta) - \frac{1}{K} \sum_{i=1}^K \log f(x_i; \theta) \quad (۸-۲)$$

محاسبه می شود، اقدام به یاد گیری پارامترهای مدل می کنیم.

در ادامه بر اساس نحوه ی تعریف تابع مدل احتمال، سه حالت مختلف را بررسی می کنیم و شرح می دهیم که الگوریتم یادگیری مقابله چیست، چرا و در چه حالتی به آن نیاز داریم.

اول حالتی را در نظر بگیرید که در آن تابع مدل احتمال^۴ که آن را با $p(x; \theta)$ نشان دادیم، تابع چگالی احتمال یک توزیع نرمال^۵ به شکل $N(x; \mu, \sigma)$ باشد. در این صورت مجموعه پارامترهای ما $\theta = \{\mu, \sigma\}$ خواهد بود. بدیهی است که انتگرال این تابع چگالی احتمال برابر با ۱ است. در نتیجه $\log Z(\theta) = 0$ می شود. همچنین

¹Data Point

²Partition Function

³Training Set

⁴Probability Model Function

⁵Normal Distribution

مشتق رابطه‌ی ۸-۲ نسبت به μ نشان می‌دهد که مقدار بهینه برای این متغیر برابر با میانگین داده‌های آموزش است. همچنین مشتق رابطه‌ی ۸-۲ این بار نسبت به σ نشان می‌دهد که مقدار بهینه برای این متغیر برابر با ریشه‌ی دوم واریانس داده‌های آموزش است.

در بعضی مواقع مانند آنچه که در اینجا ذکر کردیم، روشی وجود دارد که به صورت دقیق توانایی کمینه کردن تابع انرژی را دارد. برای درک بهتر، اگر تابع انرژی را در فضای پارامترهای مساله به صورت یک زمین موج^۱ در نظر بگیرید که هدف ما پیدا کردن پایین‌ترین نقطه در آن است، در این صورت حالتی که در اینجا ذکر شد برابر است با زمانی که در این زمین موج همه چیز واضح و هوا آفتابی است و ما پایین‌ترین نقطه را مشاهده می‌کنیم و مستقیماً به سمت آن قدم می‌زنیم [۱۱].

برای حالت بعدی زمانی را تصور کنید که در آن تابع مدل احتمال، برابر با مجموع N توزیع نرمال به شکل

$$f(x; \theta) = \sum_{i=1}^N N(x; \mu_i, \sigma_i) \quad (9-2)$$

باشد. در این حالت مجموعه پارامترهای مدل به شکل $\theta = \{\mu_1, \dots, \mu_N, \sigma_1, \dots, \sigma_N\}$ خواهد بود. این حالت مشابه زمانی است که ما یک مدل مخلوط یا مجموعه‌ای از خبره‌ها^۲ داشته باشیم که در آن وزن تمام افراد خبره برابر است. با توجه به این که انتگرال توزیع نرمال برابر با ۱ است، از رابطه ۶-۲ داریم $\log Z(\theta) = \log N$. در این حالت مشتق گرفتن از رابطه ۸-۲ نسبت به هر یک از پارامترهای مدل، مدل جدیدی که وابسته به دیگر پارامترهای مدل است را تولید می‌کند. بنابراین مقدار بهینه برای پارامترهای مدل را به صورت مستقیم نمی‌توانیم محاسبه کنیم. راه جایگزین در این حالت استفاده از معادلات مشتقات جزئی^۳ و روش نزول گرادیانی^۴ با جستجوی خطی^۵ است، که با استفاده از آن‌ها می‌توانیم نقطه کمینه محلی برای انرژی در فضای پارامترهای مدل را پیدا کنیم.

اگر به مثال خود برای حالت قبل باز گردیم، شرایط بیان شده در اینجا، یعنی استفاده از نزول گرادیانی به همراه جستجوی خطی مشابه زمانی است که ما در هنگام شب به همراه یک مشعل در آن زمین موج حضور داشته باشیم. همچنین ما توانایی احساس شیب نقطه‌ای که در آن ایستاده‌ایم را خواهیم داشت، یا می‌توانیم شیب این نقطه را نسبت به تمام جهت‌های اطرافمان تا فاصله‌ی اندکی با استفاده از نور مشعل تعیین کنیم. بنابراین با تاباندن نور مشعل به جهتی که برای حرکت انتخاب کرده‌ایم می‌توانیم پایین‌ترین نقطه در آن جهت را مشاهده کنیم به آنجا رفته و سپس جهت جدیدی را برای حرکت انتخاب کنیم [۱۱].

¹Undulating Field

²Mixture of Expert

³Partial Differential

⁴Gradient Descent

⁵Line Search

برای حالت سوم که حالت پایانی نیز است فرض کنید که ما تابع مدل احتمال را به صورت ضرب N توزیع نرمال به شکل

$$f(x; \theta) = \prod_{i=1}^N N(x; \mu_i, \sigma_i) \quad (۱۰-۲)$$

در نظر می‌گیریم. این شرایط هم ارز با یک مدل ضرب خبره‌ها است. در این حالت تابع قسمت‌بندی دیگر مقدار ثابتی نخواهد داشت و بسته به مقادیر پارامترهای توزیع‌های نرمال مقدار آن متفاوت خواهد بود. برای مثال زمانی که مدل فقط شامل دو توزیع نرمال است و برای هر دوی آن‌ها $\sigma = 1$ را در نظر بگیرید. اگر $\mu_1 = -\infty$ و $\mu_2 = \infty$ آنگاه $Z(\theta) = 0$ می‌شود. در حالی که اگر $\mu_1 = \mu_2 = 0$ آنگاه $Z(\theta) = \frac{1}{2}\sqrt{\pi}$ می‌شود. بنابراین بسته به مقادیری که برای پارامترهای مدل انتخاب می‌شوند مقدار $Z(\theta)$ متغیر خواهد بود.

اگرچه در این حالت محاسبه‌ی دقیق تابع قسمت‌بندی امکان‌پذیر است، اما شرایطی را در نظر بگیرید که در آن تابع مدل احتمال به گونه‌ای باشد که محاسبه‌ی انتگرال در رابطه ۲-۶ از نظر جبری غیرعملی^۱ باشد. در این حالت برای ارزیابی رابطه ۲-۸ ما نیاز خواهیم داشت از انتگرال‌گیری عددی^۲ استفاده کنیم. همچنین باید از مشتقات متناهی برای محاسبه گرادیان در یک نقطه‌ی داده شده در فضای پارامترهای مساله و همچنین روش‌های نزول گرادیانی برای پیدا کردن کمینه محلی استفاده کنیم. برای فضاهای داده‌ای و پارامتری با ابعاد بالا زمان این انتگرال‌گیری فلج‌کننده^۳ خواهد بود. شرایط بیان شده تا اینجا منجر به وضعیتی می‌گردد که در آن ما تلاش می‌کنیم یک تابع انرژی را کمینه کنیم در حالی که توانایی ارزیابی کردن آن را نداریم.

اینجا زمانی است که الگوریتم واگرایی مقابله به ما کمک می‌کند. اگر چه که ما نمی‌توانیم خود تابع انرژی را ارزیابی کنیم اما الگوریتم CD برای ما راهی را مهیا می‌کند که می‌توانیم به کمک آن گرادیان تابع انرژی را تخمین بزنیم. اگر به مثال خود برای حالت‌های قبل بازگردیم، در شرایط بیان شده برای حالت نهایی ما خودمان را در همان زمین موج بدون هیچ‌گونه امکانات و توانایی خاصی خواهیم یافت (ما نمی‌توانیم انرژی را محاسبه کنیم)، در نتیجه ما امکان تشخیص ارتفاع و شیب در هیچ یک از نقطه‌های اطرافمان را نسبت به نقطه‌ای که در آن ایستاده‌ام را نداریم. الگوریتم CD در این حالت به ما یک حس تعادل می‌دهد و به ما این اجازه را می‌دهد که شیب نقطه‌ای از زمین که در زیر پایمان قرار دارد را تشخیص دهیم. حال با برداشتن قدم‌های بسیار کوچک در جهتی که بیشترین کاهش شیب را داریم ما توانایی پیدا کردن راه خودمان به سمت کمینه‌ی محلی را خواهیم داشت [۱۱].

¹Intractable

²Numerical Integration

³Crippling

۲-۱۰-۲ الگوریتم واگرایی مقابله چگونه کار می کند

همان طور که در بخش ۲-۱۰-۱ توضیح داده شد، الگوریتم CD با توجه مجموعه پارامترهای مدل و داده های آموزش، گرادیان تابع انرژی را برای ما تخمین می زند. در رابطه ی

$$\begin{aligned}\frac{\partial E(\mathbf{X}; \theta)}{\partial \theta} &= \frac{\log Z(\theta)}{\partial \theta} - \frac{1}{K} \sum_{i=1}^K \frac{\partial \log f(x_i; \theta)}{\partial \theta} \\ &= \frac{\log Z(\theta)}{\partial \theta} - \left\langle \frac{\partial \log f(x_i; \theta)}{\partial \theta} \right\rangle_{\mathbf{X}}\end{aligned}\quad (11-2)$$

با استفاده از مشتقات جزئی فرمول محاسبه ی گرادیان را از رابطه ۲-۸ بدست می آوریم. در رابطه ی ۲-۱۱ نماد $\langle * \rangle_{\mathbf{X}}$ نشان دهنده مقدار مورد انتظار برای $*$ با توجه به توزیع داده ی \mathbf{X} می باشد.

اولین ترم در سمت راست رابطه ی ۲-۱۱ از تابع قسمت بندی مشتق می گردد و همانطور که در رابطه ی ۲-۶ مشاهده می گردد شامل یک انتگرال گیری بر روی x است. با جایگذاری این رابطه به رابطه ی

$$\begin{aligned}\frac{\partial \log Z(\theta)}{\partial \theta} &= \frac{1}{Z(\theta)} \frac{\partial Z(\theta)}{\partial \theta} \\ &= \frac{1}{Z(\theta)} \frac{\partial}{\partial \theta} \int f(x; \theta) dx \\ &= \frac{1}{Z(\theta)} \int \frac{\partial f(x; \theta)}{\partial \theta} dx \\ &= \frac{1}{Z(\theta)} \int f(x; \theta) \frac{\partial \log f(x; \theta)}{\partial \theta} dx \\ &= \int p(x; \theta) \frac{\partial \log f(x; \theta)}{\partial \theta} dx \\ &= \left\langle \frac{\partial \log f(x; \theta)}{\partial \theta} \right\rangle_{p(x; \theta)}\end{aligned}\quad (12-2)$$

می رسم. همان طور که در بخش ۲-۱۰-۱ بحث کردیم محاسبه ی این انتگرال در حالت کلی از نظر جبری غیر عملی است. اما در رابطه ی ۲-۱۲ روشن است که می توانیم با نمونه گرفتن از توزیع $p(x; \theta)$ این انتگرال را به صورت عددی تقریب بزنیم. از آن جا که ما مقدار دقیق تابع قسمت بندی را نمی دانیم لذا نمی توانیم به صورت مستقیم از $p(x; \theta)$ نمونه تولید کنیم. اما با استفاده از تعداد زیادی چرخه نمونه برداری MCMC که در بخش ۲-۹ معرفی شد می توانیم توزیع داده های آموزش را به توزیع پیشنهاد شده برای داده ها تبدیل کنیم. این تبدیل از آنجایی امکان پذیر است که تنها شامل محاسبه ی نسبت دو توزیع احتمالی به صورت $\frac{p(x'; \theta)}{p(x; \theta)}$ است و تابع قسمت بندی از معادله حذف می گردد. در اینجا \mathbf{X}^n نشان دهنده ی داده های آموزش است که با n چرخه ی نمونه برداری MCMC تبدیل یافته اند، با توجه به این می توانیم تعریف کنیم: $\mathbf{X}^0 = \mathbf{X}$. با جایگذاری این روابط

در رابطه‌ی ۲-۱۱ به رابطه‌ی

$$\frac{\partial E(\mathbf{X}; \theta)}{\partial \theta} = \left\langle \frac{\partial \log f(x; \theta)}{\partial \theta} \right\rangle_{\mathbf{X}^\infty} - \left\langle \frac{\partial \log f(x; \theta)}{\partial \theta} \right\rangle_{\mathbf{X}^0} \quad (۱۳-۲)$$

می‌رسیم.

تعداد زیاد چرخه‌ی نمونه‌برداری MCMC برای محاسبه‌ی یک گرادیان صحیح تنها مانع محاسباتی باقی مانده است که باید بر آن غلبه کنیم. هینتون بیان کرد که تنها تعداد کمی چرخه‌ی نمونه‌برداری MCMC برای محاسبه‌ی یک گرادیان تقریبی نیاز خواهد بود، و بعد از تعداد کمی تکرار توزیع داده‌های آموزش به سمت توزیع پیشنهاد شده شروع به حرکت می‌کند و جهتی را که در آن داده‌ها بهتر مدل می‌شوند را نشان می‌دهد [۱۰][۱۲]. پیاده‌سازی‌های عملی نشان داد که تنها یک چرخه‌ی نمونه‌برداری MCMC برای همگرایی الگوریتم به مقدار بیشینه مورد انتظار کفایت می‌کند [۱۲]. در نتیجه به منظور کمینه کردن تابع انرژی رابطه‌ی به روزرسانی کردن مجموعه پارامترهای ما به شکل رابطه‌ی

$$\theta_{t+1} = \theta_t + \eta \left(\left\langle \frac{\partial \log f(x; \theta)}{\partial \theta} \right\rangle_{\mathbf{X}^0} - \left\langle \frac{\partial \log f(x; \theta)}{\partial \theta} \right\rangle_{\mathbf{X}^1} \right) \quad (۱۴-۲)$$

می‌گردد که در آن η ضریب یادگیری است و باید بر اساس زمان همگرایی و ثبات یادگیری الگوریتم به صورت تجربی مشخص گردد.

۲-۱۱ کیسه‌ی کلمات

مفهوم کیسه‌ی کلمات^۱ بیشتر در مباحث مربوط به پردازش زبان طبیعی و بازیابی اطلاعات مورد استفاده قرار می‌گیرد و از آن به عنوان راهی برای نمایش داده استفاده می‌کنند. در این روش هر جمله و یا هر سند به صورت کیسه‌ای از کلماتش که در واقع برداری از اعداد صحیح هستند نشان داده می‌شود. در این بردار ترتیب کلمات و اصول و قواعد گرامری رعایت نمی‌شود و تنها شامل تعداد تکرار کلمات متمایز و مختلف در جمله یا سند متناظرش است. تعداد تکرار هر کلمه در روش کیسه کلمات به عنوان یک ویژگی برای هر سند یا جمله در نظر گرفته می‌شود.

برای مثال دو متن زیر که به ترتیب دارای دو و یک جمله می‌باشند را در نظر بگیرید:

1. John likes to watch movies. Mary likes movies too.
2. John also likes to watch football games.

برای نشان دادن این دو سند با استفاده از روش کیسه‌ی کلمات اینگونه عمل می‌کنیم:

۱. شروع به پیمایش متن‌ها کرده و تمام کلمات متمایز را مشخص می‌کنیم.

”John”, ”likes”, ”to”, ”watch”, ”movies”, ”also”, ”football”, ”games”, ”Mary”, ”too”

^۱ Bag of Words

۲. کلمات متمایز را شماره‌گذاری می‌کنیم. در واقع یک لغت‌نامه تشکیل می‌دهیم که در آن هر عدد متناسب با یک لغت متمایز است.

John=1, likes=2, to=3, watch=4, movies=5, also=6, football=7, games=8, Mary=9, too=10

۳. تمام اسناد را با برداری به طول ثابت نشان می‌دهیم. اندازه‌ی این بردار برابر با بزرگترین اندیس کلمات در لغت‌نامه‌ی ساخته شده در مرحله‌ی قبل است، و هر درایه‌ی آن متناظر با لغتی با همان اندیس در لغات‌نامه‌ی ساخته شده است.

۴. برای هر سند بردار متناظر با آن را تشکیل می‌دهیم و تعداد تکرار کلمات مختلف لغت‌نامه را در بردار نشان می‌دهیم.

(a) {1, 2, 1, 1, 2, 0, 0, 0, 1, 1}

(b) {1, 1, 1, 1, 0, 1, 1, 1, 0, 0}

۱۲-۲ موضوع

در حوزه‌ی پردازش زبان طبیعی یک موضوع به مجموعه یا خوشه‌ای^۱ از کلمات متمایز گفت می‌شود که از نظر معنایی به یکدیگر نزدیک هستند و در یک دسته قرار می‌گیرند. برای مثال در شکل ۱-۲ چهار خوشه از کلمات را مشاهده می‌کنید که از نظر معنایی به هم نزدیک هستند و می‌توانیم به ترتیب موضوع‌های الکترونیکی، فضایی، ورزشی و مذهبی را به آن‌ها اختصاص دهیم.

card	shuttle	team	christianity
driver	orbit	games	god
drivers	lunar	seasons	pgp
bus	spacecraft	baseball	jesus
video	nasa	players	bible
vga	space	game	faith
monitor	launch	hockey	muslim
ibm	saturn	play	christ
cards	billion	teams	atheist
ram	satellite	sale	christians

جدول ۱-۲: نمونه‌ی چهار موضوع مختلف

۱۳-۲ مدل موضوعی

مدل‌های موضوعی یک کلاس جدید از روش‌های آنالیز متن^۲ هستند که به تازگی به صورت گسترده مورد توجه پژوهشگران حوزه‌های مختلف مانند پردازش زبان طبیعی، علوم اجتماعی، اقتصاد و غیره قرار گرفته‌اند

^۱Cluster

^۲Text Analysis

[۱۳][۱۴]. در مباحث مربوط به یادگیری ماشین و پردازش زبان طبیعی یک مدل موضوعی یک دسته از مدل‌های آماری است که برای کشف و نمایش یک چکیده^۱ از موضوع‌هایی که در یک مجموعه از اسناد^۲ وجود دارد به کار می‌رود [۸][۱۵]. مهمترین ویژگی این مدل‌ها که باعث تمایز آن‌ها از دیگر مدل‌ها می‌گردد، روند کاملاً اتوماتیک و خودکار آن‌ها برای تبدیل محتوای یک مجموعه سند متنی که این مجموعه سند می‌تواند بسیار بزرگ باشد به دسته‌هایی با معنا و مفهوم که "موضوع‌ها" نامیده می‌شوند، است [۱۳].

الگوریتم‌ها و روش‌های موجود در این حوزه این روند را با کمترین دخالت انسانی و به صورت کاملاً خودکار انجام می‌دهند و همین امر موجب می‌گردد که به صورت قابل ملاحظه‌ای به روش‌های پیشین موجود در این زمینه ترجیح داده شوند. چرا که بیشتر روش‌های موجود در این زمینه نیاز به دخالت و نظارت یک عامل انسانی در تمام مراحل اجرا را دارند. در مدل‌های موضوعی هیچ دانش اولیه‌ای در رابطه با نوع دسته‌ها یا موضوع‌های موجود در مجموعه سند داده نمی‌شود و تنها تعداد موضوع‌های مورد نظر به مدل داده می‌شود و پس از اتمام روند مدل‌سازی این موضوع‌ها به صورت توزیع‌های احتمالاتی بر روی کلمات متمایز موجود در مجموعه سند مشخص می‌شوند [۱۳].

اما بحث مدل‌های موضوعی تنها به داده‌های متنی خلاصه نمی‌شود، به طور مثال در مباحث مربوط به پردازش تصویر^۳ و استخراج ویژگی^۴ از تصویر و یا ارایه‌ی تصویر به صورت فشرده می‌توان در ابتدا با استفاده از الگوریتم‌ها و روش‌های موجود در زمینه‌ی پردازش تصویر به تبدیل هر تصویر به یک سند متنی پرداخت و سپس با اجرای مدل‌های موضوعی بر روی سند متنی بدست آمده نتیجه‌ی مورد نیاز را به خوبی بدست آورد [۱۳].

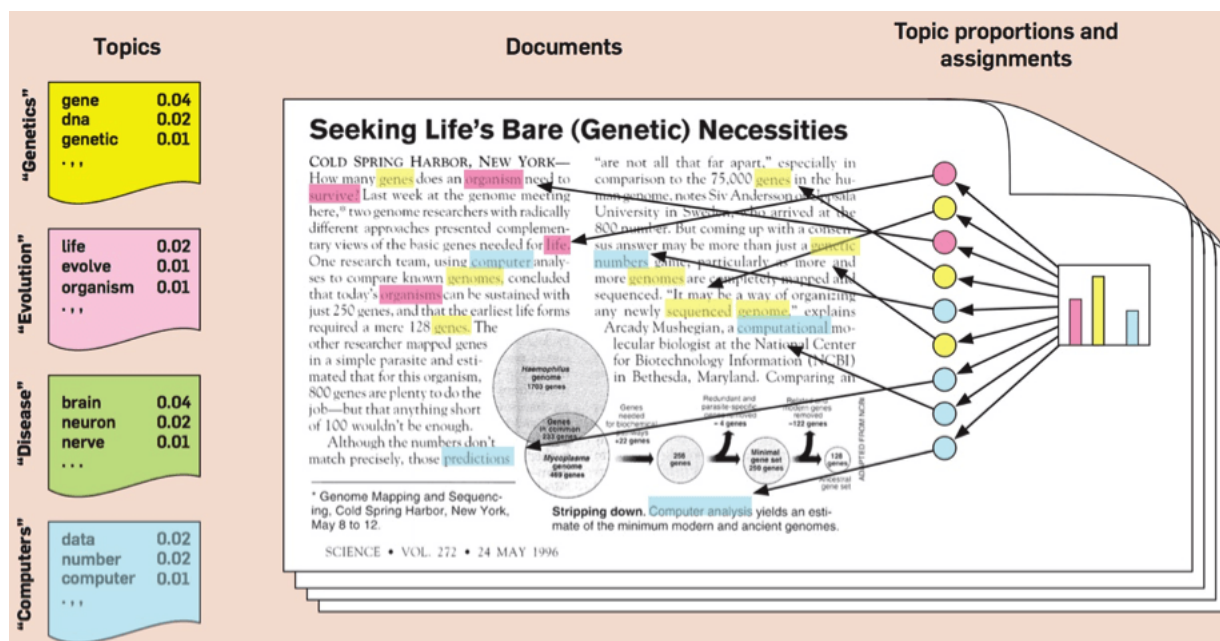
اینکه این مدل‌ها به چه صورت عمل می‌کنند و بر چه اساسی موضوع‌های موجود در یک مجموع سند را برای ما مدل می‌کنند یا به صورت کلی در پاسخ به این سوال که تئوری و ایده‌ی اصلی برای این مدل‌ها کدام است را می‌توان در یک پاسخ ساده خلاصه کرد. تمام این مدل‌ها بر اساس این فرضیه ساخته می‌شوند که معانی و مفاهیم با یکدیگر مرتبط هستند. به بیان دیگر به طور مثال در صحبت‌های روزمره مفاهیم و کلماتی که به صورت ذاتی در مورد یک موضوع خاص هستند یک مجموعه یا خوشه از کلمات به خصوصی را شامل می‌شوند. بنابراین یک موضوع خاص می‌تواند به صورت مجموعه‌ای از کلماتی در نظر گرفته شود که از نظر معنایی به هم نزدیک هستند و تمایل دارند که در یک مکالمه یا یک نوشتار زمانی که در مورد آن موضوع خاص بحث می‌شود در کنار هم تکرار شوند و رخ دهند. توجه به این نکته ضروری است که از آنجا که مدل‌های موضوعی با کیسه‌ی کلمات کار می‌کنند در نتیجه تکرار و رخداد کلمات در کنار یکدیگر را به عنوان ویژگی مورد توجه قرار

¹ Abstract

² Corpus

³ Image Processing

⁴ Feature Extraction



شکل ۲-۳: نمونه‌ای از یک مدل‌سازی موضوعی

می‌دهند و از پیچیدگی‌های دیگر زبانی مانند قواعد گرامری، دستوری، نحوی و غیره صرف نظر می‌کنند. اگرچه در تعدادی از مدل‌های موجود در این زمینه به این پیچیدگی‌های زبانی نیز اندکی توجه گردیده است اما در بیشتر این مدل‌ها از این پیچیدگی‌ها صرف نظر می‌شود. در نتیجه هدف نهایی از مدل‌سازی موضوعی آنالیز خوشه‌های مختلف کلمات و تشخیص کلماتی است که از نظر معنایی به هم نزدیک هستند و تمایل دارند در کنار یکدیگر تکرار شوند و یک موضوع خاص را شکل دهند [۱۳]. در مدل‌سازی موضوعی فرض بر این است که هر سند شامل چند موضوع مختلف است که هرکدام از این موضوع‌ها به صورت یک توزیع احتمالاتی بر روی کلمات بخصوصی هستند [۱۶].

برای مثال در شکل ۲-۳ نمونه‌ای از یک مدل‌سازی موضوعی را مشاهده می‌کنید. در این شکل چهار موضوع با رنگ‌های مختلف به همراه احتمال رخ داد کلمات مرتبط با آن نشان داده شده‌اند. همچنین می‌توان نحوی تکرار این کلمات در متن اصلی را مشاهده کرد که هر کلمه با رنگ مرتبط با یک موضوع مشخص شده است.

۲-۱۴ آنالیز احساس

منظور از آنالیز احساس یا کاوش عقاید مشخص کردن اطلاعات مفهومی^۱ مانند نظرات، نگرش‌ها و احساس موجود در متن نوشته شده است. در آنالیز احساس در حوزه‌ی پردازش زبان طبیعی و پردازش متن ما به دنبال ابزارهایی هستیم که این اطلاعات مفهومی را به صورت خود کار از متن یا مجموعه اسناد استخراج کند و احساس

^۱ Subjective Information

موجود در متن را برای ما مشخص کند [۶][۹]. می‌توان احساس موجود در یک متن را در سه دسته قرار داد: مثبت، منفی و بی‌طرف.

هدف نهایی در آنالیز احساس مشخص کردن همین امر است که یک متن از نظر احساسی مثبت، منفی و یا بی‌طرف است. در بحث تشخیص همزمان موضوعی و احساس از یک متن فرض بر این است که در یک سند متنی در مورد چندین موضوع صحبت شده است و هرکدام از این موضوع‌ها می‌توانند احساس خاص مربوط به خود را داشته باشند. به طور مثال فرض کنید که شما یک محصول الکترونیکی جدید مانند لپ‌تاپ خریداری کرده‌اید. حال اگر شما یک بازبینی^۱ برای این لپ‌تاپ بنویسید، این نوشته شامل موضوعات مختلفی خواهد بود چرا که یک لپ‌تاپ شامل بخش‌های مختلفی نظیر صفحه نمایش، باتری، حافظه و ... است و در بازبینی نوشته شده هر کدام از این‌ها به عنوان یک موضوع مجزا در نظر گرفته می‌شوند. حال هرکدام از این موضوع‌ها می‌توانند احساس خاص مربوط به خود را داشته باشند، به طور مثال شما از حافظه لپ‌تاپ راضی هستید ولی در مقابل صفحه نمایش آن انتظارات شما را برآورده نکرده است، در اینجا احساس شما به موضوع حافظه مثبت و به موضوع صفحه نمایش منفی خواهد بود. هدف ما در تشخیص همزمان موضوع و احساس از داده‌های متنی نیز به همین شکل است که علاوه بر تشخیص موضوع‌های مورد بحث، احساس همراه با هرکدام را نیز تشخیص دهیم.

^۱ Review

فصل سوم

بررسی پژوهش‌های پیشین

۳-۱ مقدمه

در این فصل به مرور و بررسی روش‌ها و پژوهش‌های موجود پیشین در زمینه‌ی تخمین توزیع^۲، استخراج اطلاعات^۳ از متن، بازیابی اطلاعات و مدل‌سازی موضوع می‌پردازیم. در همین راستا به بررسی مهم‌ترین مدل‌ها و روش‌های موجود در زمینه‌ی مدل‌سازی موضوع، مدل‌سازی موضوع و احساس به صورت همزمان و همچنین دسته‌ی خاصی از مدل‌های موضوعی که به آن‌ها چندحالت گفته می‌شود می‌پردازیم. هدف از معرفی این روش‌ها آشنایی با کارهای انجام شده در زمینه‌ی مربوطه تا به امروز و درک هرچه بهتر ایده و علت انجام این پژوهش است. در این فصل روش‌های موجود را از چندین زاویه مورد نقد و بررسی قرار می‌دهیم و بسته به ساختار، نحوه‌ی عملکرد، نوع داده‌ی ورودی و سیر تکاملی، آن‌ها را در چندین کلاس طبقه‌بندی می‌کنیم.

در بررسی روش‌ها و مدل‌های موجود از دید ساختاری، می‌توان آن‌ها را در دو دسته کلی طبقه‌بندی کرد. یک مدل‌های بر پایه‌ی شبکه‌های عصبی و دو مدل‌های گرافی. در بحث مدل‌سازی موضوع، مدل‌های گرافی نسبت به مدل‌های شبکه‌های عصبی از قدمت بیشتری برخوردار هستند. مهم‌ترین مدلی که در این دسته وجود دارد و در بخش ۳-۳ آن را دقیق‌تر بررسی می‌نماییم مدل معروف تخصیص دیریکله‌ی پنهان است (LDA) که در

^۲Distribution Estimation

^۳Information Extraction

	Graphical Model	Neural Network Model
Latent Semantic Indexing(LSI)	*	
probabilistic Latent Semantic Indexing(pLSI)	*	
Latent Dirichlet Allocation(LDA)	*	
Joint Sentiment/Topic Model Detection(JST)	*	
Aspect and Sentiment Unification Model(ASUM)	*	
Supervised Joint Aspect and Sentiment Model(SJASM)	*	
Neural Autoregressive Distribution Estimator(NADE)		*
Restricted Boltzman Machine(RBM)		*
Replicated Softmax(RS)		*
Document NADE(DocNADE)		*
Supervised DocNADE(SupDocNADE)		*

جدول ۳-۱: دسته‌بندی مدل‌های پیشین از نظر ساختار

سال ۲۰۰۳ توسط Blei و همکاران [۱] ارائه گردید، و پس آن به عنوان پایه‌ی مدل‌سازی موضوعی در بخش مدل‌های گرافی قرار گرفت. دسته‌ی دیگر مدل‌های موضوعی موجود از نظر ساختار آن‌هایی هستند که بر پایه‌ی شبکه‌های عصبی مصنوعی هستند و اولین بار در سال ۲۰۰۹ توسط Hinton و Salakhutdinov [۳] معرفی شدند. در بخش ۳-۳ با هر کدام از این دو دسته بیشتر آشنا می‌شویم و روش‌های موجود در آن‌ها را معرفی می‌کنیم. جدول ۳-۱ ساختار تمام روش‌های معرفی شده در این فصل که در ادامه به صورت مفصل به هر کدام می‌پردازیم را نشان می‌دهد.

از نظر نحوه‌ی عملکرد مدل‌های پیشین را در سه کلاس مختلف قرار می‌دهیم. دسته‌ی اول روش‌هایی که تنها به مدل‌سازی موضوع می‌پردازند و آن‌ها را به عنوان روش‌های مدل‌سازی موضوعی معرفی می‌کنیم. دسته‌ی دوم روش‌هایی که تنها به تشخیص احساس و دانش مفهومی از داده‌های ورودی می‌پردازند. اگرچه باید توجه داشت که مدل‌های موجود در زمینه تشخیص احساس در دسته‌ی مدل‌های موضوعی قرار نمی‌گیرند و بیشتر شامل مدل‌های یادگیری ماشین هستند که یک طبقه‌بندی^۱ دو حالتی (مثبت و منفی) یا سه حالتی (مثبت و منفی، مثبت و بی طرف) را انجام می‌دهند. و در دسته‌ی سوم که دسته‌ی آخر از نظر نحوه‌ی عملکرد است روش‌هایی را بررسی می‌کنیم که به صورت همزمان به مدل‌سازی موضوع و احساس بر روی داده‌ی ورودی می‌پردازند. در جدول ۳-۲ نوع مدل‌های معرفی شده در این فصل از نظر نحوه‌ی عملکرد مشاهده می‌شود.

روش‌های پیشین از نظر نوع داده‌ی ورودی در دو کلاس متفاوت قرار می‌گیرند. یک گروه روش‌هایی که تنها یک نوع داده را به عنوان ورودی قبول می‌کنند. منظور از یک مدل داده این است که روش‌های موجود توانایی عمل کردن به صورت همزمان بر روی چند مد مختلف از داده‌ها را ندارند، و داده‌های ورودی تنها باید یک حالت داشته باشند، مثلاً تنها متن و یا تنها تصویر باشند و نمی‌توانند ترکیبی از این‌ها باشند. دسته‌ی دوم که آن‌ها را مدل‌های چندحالتی می‌شناسیم مدل‌هایی هستند که با داده‌های چندوجهی کار می‌کنند. منظور از داده‌های

^۱ Classification

	Topic Model	Sentiment/Topic Model
Latent Semantic Indexing(LSI)	*	
probabilistic Latent Semantic Indexing(pLSI)	*	
Latent Dirichlet Allocation(LDA)	*	
Joint Sentiment/Topic Model Detection(JST)		*
Aspect and Sentiment Unification Model(ASUM)		*
Supervised Joint Aspect and Sentiment Model(SJASM)		*
Replicated Softmax(RS)	*	
Document NADE(DocNADE)	*	

جدول ۳-۲: دسته‌بندی مدل‌های پیشین از نظر نحوه‌ی عملکرد

چندوجهی آن‌هایی هستند که شامل ترکیب چند حالت مختلف از داده هستند، برای مثال ترکیب متن و تصویر و یا ترکیب تصویر و صدا. در جدول ۳-۲ تفاوت بین روش‌های موجود از نظر نوع داده‌ی ورودی مشخص شده است.

از نقطه‌نظر سیر تکاملی می‌توان روش‌های موجود را در سه سطح: یک مدل‌های تخمین زنده‌ی توزیع، دو روش‌های مدل‌سازی موضوع و سه مدل‌های تشخیص همزمان موضوع و احساس به صورت مشترک قرار داد. البته لازم به ذکر است که روش‌های تخمین توزیع که در اینجا مطرح می‌گردند و در بحث پردازش زبان طبیعی مورد استفاده قرار می‌گیرند به تنهایی در دسته‌ی مدل‌های موضوعی قرار نمی‌گیرند اما پایه و اساس بسیاری از مدل‌های موضوعی هستند و لذا در بخش ۳-۲ آن‌ها را معرفی کرده و به صورت مختصر توضیح می‌دهیم.

۳-۲ مدل‌های تخمین توزیع احتمالی

اولین گروه از مدل‌های پیشین که به بررسی آن‌ها می‌پردازیم روش‌هایی هستند که به تخمین توزیع‌های احتمالی موجود در داده‌های ورودی می‌پردازند. در بخش‌های ۳-۲-۱ و ۳-۲-۲ به ترتیب مدل‌های ”ماشین بلترن

	Unimodal	Multimodal
Latent Semantic Indexing(LSI)	*	
probabilistic Latent Semantic Indexing(pLSI)	*	
Latent Dirichlet Allocation(LDA)	*	
Joint Sentiment/Topic Model Detection(JST)	*	
Aspect and Sentiment Unification Model(ASUM)	*	
Supervised Joint Aspect and Sentiment Model(SJASM)	*	
Neural Autoregressive Distribution Estimator(NADE)	*	
Restricted Boltzman Machine(RBM)	*	
Replicated Softmax(RS)	*	
Document NADE(DocNADE)	*	
Supervised DocNADE(SupDocNADE)		*

جدول ۳-۳: دسته‌بندی مدل‌های پیشین از نظر داده‌ی ورودی

محدود^۱ (RBM) [۱۷][۱۸][۱۰] و همچنین مدل "شبکه‌ی عصبی خود رگرسیو تخمین‌زننده‌ی توزیع"^۲ (NADE) که در سال ۲۰۱۱ توسط Larochelle و Murray [۲] معرفی شد، را بررسی می‌کنیم. این دو مدل از مهمترین روش‌های موجود در زمینه تخمین توزیع در بحث پردازش زبان طبیعی هستند. دلیل اهمیت و معرفی این روش‌ها این می‌باشد که مدل‌هایی که در بخش‌های بعدی معرفی می‌شوند گسترش یافته‌ی این روش‌ها هستند و از تغییر در ساختار این مدل‌ها بدست می‌آیند.

۳-۲-۱ مدل ماشین بلتزن محدود

مدل ماشین بلتزن محدود که به اختصار آن را RBM می‌نامیم یک مدل شبکه عصبی بدون نظارت برای تخمین توزیع داده‌های ورودی باینری است. RBM در دسته مدل‌های احتمالاتی مولد (بخش ۲-۳-۲) قرار می‌گیرد که می‌تواند توزیع احتمالی داده‌های ورودی خود را یاد بگیرد. شبکه RBM اولین بار در سال ۱۹۸۶ توسط Smolensky [۱۷] و پس از آن در سال ۲۰۰۲ به شکل دیگری توسط Hinton [۱۰] معرفی گردید. برای آموزش این شبکه از الگوریتم CD که در بخش ۲-۱۰ توضیح داده شد استفاده می‌شود.

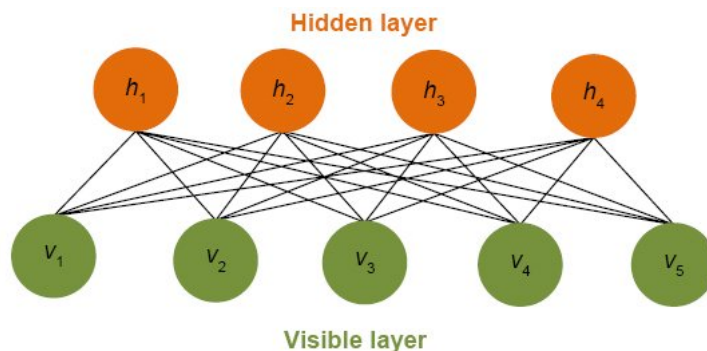
RBM از یک ساختار دو لایه، یک لایه‌ی قابل مشاهده و یک لایه‌ی پنهان، مانند شکل ۳-۱ تشکیل شده است. از نظر ریاضی شبکه‌ی RBM ساختاری مانند یک گراف دوبخشی دارد، به عبارت دیگر هر نورون در لایه‌ی قابل مشاهده به تمام نورون‌ها در لایه‌ی مخفی متصل است و بر عکس، و همچنین در داخل هر لایه هیچ اتصالی بین نورون‌ها وجود ندارد.

RBM یک مدل بر پایه‌ی انرژی است و هدف نهایی در آن بدست آوردن پارامترهای مدل به گونه‌ای است که به ازای هر بردار داده‌ی ورودی و لایه‌ی پنهان متناسب با آن، مدل در پایین‌ترین سطح انرژی خود باشد [۱۹]. لایه‌ی پنهان در این مدل نیز همانند لایه‌ی ورودی دارای ساختار باینری است و واحدهای آن می‌توانند به صورت احتمالاتی ۰ یا ۱ باشند. در مدل RBM لایه‌ی پنهان را می‌توان هم‌ارز با یک بردار ویژگی که از بردار ورودی بدست می‌آید دانست، به این صورت که هر بردار باینری از داده‌های ورودی بردار پنهان مخصوص به خود را خواهد داشت.

محدود بودن به طول بردار ثابت برای داده‌های ورودی و همچنین حالت باینری برای آن‌ها از مشکلات مدل RBM است که علی‌رغم قدرت بسیار بالای این مدل در تخمین توزیع داده‌های ورودی باعث گردیده در کاربردهای دنیای واقعی آنچنان که باید مورد استفاده قرار نگیرد.

¹Restricted Boltzman Machine

²Neural Autoregressive Distribution Estimator



شکل ۳-۱: ماشین بلترمن محدود

۳-۲-۲ مدل شبکه‌ی عصبی خود رگرسیو تخمین‌زننده‌ی توزیع

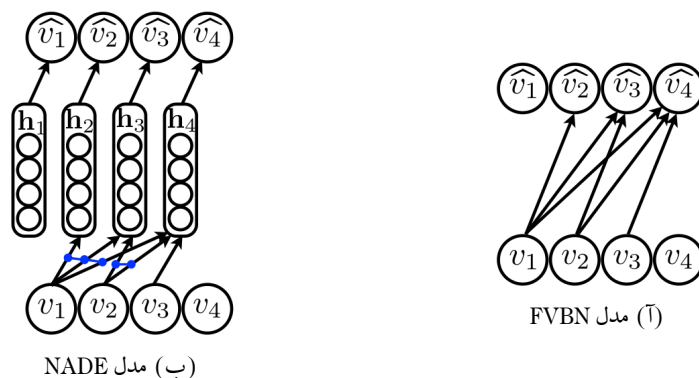
مدل NADE که از مدل RBM الهام گرفته شده است، یک روش احتمالاتی مولد بدون نظارت برای مدل‌سازی احتمال داده‌های گسسته در ابعاد بالا است [۲]. در این روش نیز همانند مدل RBM طول بردار ورودی ثابت در نظر گرفته می‌شود، همچنین ورودی نیز مانند شبکه‌ی RBM محدود به حالت باینری است [۲]. مشکل دیگر موجود در مدل RBM مناسب نبودن این مدل برای تخمین احتمال مشترک در ابعاد بالا است. این مشکل در مدل NADE بدلیل استفاده از ایده‌ی شبکه‌های بیزین کاملاً مشاهده‌پذیر^۱ (FVBN) [۲۰][۲۱] برای محاسبه‌ی احتمال مرتفع گردیده است. در این شبکه‌ها برای محاسبه‌ی احتمال مشاهده‌ی یک بردار ورودی مانند \mathbf{V} توسط مدل از رابطه‌ای مشابه رابطه‌ی

$$p(\mathbf{V}) = \prod_{i=1}^D (v_i | \mathbf{V}_{parent(i)}) \quad (۳-۱)$$

استفاده می‌شود، که در آن $\mathbf{V}_{parent(i)} = \mathbf{V}_{<i}$ اشاره به یک زیربردار دارد که شامل تمام ابعاد قبل از i امین بعد است. به بیان دیگر در این شبکه‌ها قبل از ساختن مدل، برای داده‌های ورودی یک ترتیب فرضی برای ابعاد آن در نظر گرفته می‌شود و سپس یک بردار جهت‌دار به مدل به عنوان ورودی داده می‌شود و $\mathbf{V}_{<i}$ یک بردار جهت‌دار شامل تمام ابعاد تا قبل از i امین بعد است. با تعریف بیان شده می‌توان گفت که در این شبکه‌ها احتمال هر بعد از بردار ورودی مشروط به تمام ابعاد قبل از آن محاسبه می‌گردد. بنابراین اگر تمام $p(v_i | \mathbf{V}_{parent(i)})$ ها از نظر جبری قابل محاسبه باشند در نتیجه بدست آوردن مقدار $p(\mathbf{V})$ از نظر محاسباتی امکان پذیر است [۲].

در مدل NADE نیز که از رابطه‌ی ۳-۱ برای محاسبه‌ی احتمال یک بردار ورودی استفاده می‌شود، می‌توان احتمال داده‌های با ابعاد بالا را مشروط به تمام ابعادش محاسبه کرد. در شکل‌های ۳-۲آ و ۳-۲ب به ترتیب مدل‌های FVBN و NADE نشان داده شده‌اند. همان‌طور که بیان شد و در شکل ۳-۲ب نیز مشاهده می‌گردد در مدل NADE به طور مثال مقدار احتمال بعد سوم بردار ورودی که با \hat{v}_3 نشان داده می‌شود از یک لایه‌ی پنهان

^۱Fully Visible Bayesian Networks



شکل ۳-۲: مدل‌های FVBN و NADE [۲]

بدست می‌آید که تنها وابسته به ابعاد قبل از v_3 یعنی v_1 و v_2 در بردار ورودی است [۲].

۳-۳ روش‌های مدل‌سازی موضوعی

همان‌طور که پیش از این در بخش ۲-۱۳ بیان کردیم روش‌های مدل‌سازی موضوع به مدل‌هایی گفته می‌شود که یک چکیده از موضوعات موجود در یک سند یا مجموعه‌ای از اسناد را تشخیص داده و استخراج می‌کنند. این مدل‌ها را می‌توان در دو دسته کلی شامل مدل‌های گرافی که بر پایه‌ی قوانین احتمالاتی و رابطه‌ی بیز هستند و همچنین مدل‌های شبکه‌های عصبی تقسیم‌بندی کرد. در ادامه ضمن معرفی هر کدام از این کلاس‌ها با مهمترین مدل‌های موجود در این حوزه آشنا می‌گردیم.

۱-۳-۳ مدل تکرار عبارت-معکوس تکرار سند

تاکنون محققین حوزه‌ی بازیابی اطلاعات پیشرفت‌های قابل توجهی در زمینه‌ی مدل کردن مجموعه‌ی اسناد یا هر مجموعه‌ی گسسته از داده‌ها داشته‌اند [۲۲]. در روش پایه که توسط پژوهشگران حوزه‌ی IR پیشنهاد گردیده و امروزه همچنان در مرورگرهای اینترنتی مورد استفاده قرار می‌گیرد، هر سند متنی از یک مجموعه اسناد به یک بردار اعداد حقیقی تبدیل می‌شود که شامل نسبت‌های تعداد تکرار کلمات مختلف است.

در این روش که به آن تکرار عبارت-معکوس تکرار سند^۱ ($tf-idf$) [۲۳] گفته می‌شود و در سال ۱۹۸۶ توسط Salton معرفی شد یک لغات‌نامه از تمام کلمات متمایز ساخته می‌شود، سپس برای هر سند در مجموعه‌ی اسناد تعداد رخداد تمام کلمات متمایز محاسبه می‌شود. پس از نرمال‌سازی مناسب (بیشتر در مقیاس لگاریتمی) مقادیر بدست آمده برای هر کلمه، هر مقدار بر معکوس تعداد سندهای شامل آن کلمه در کل مجموعه‌ی اسناد تقسیم می‌گردد. مقادیر نهایی بدست آمده برای هر کلمه که به آن مقادیر $tf-idf$ گفته می‌شود به صورت یک بردار ستونی در یک ماتریس جایگذاری می‌شوند. بنابراین در روش $tf-idf$ یک مجموعه سند به ماتریسی

^۱ Term Frequency-Inverse Document Frequency

$m \times n$ تبدیل می‌شود که در آن m تعداد کلمات متمایز در مجموعه سند و n تعداد سندهای موجود در مجموعه سند هستند.

به عنوان مثال یک مجموعه اسناد که تنها از دو سند تشکیل شده است را در نظر بگیرید. اگر جدول تکرار کلمه‌ها برای هر سند مانند شکل ۳-۳ باشد، آنگاه محاسبه‌ی $tf-idf$ برای یک کلمه مانند "this" به صورت زیر است:

$$\begin{aligned} Document_1 : tf("this", d_1) &= \frac{1}{5} = 0.2 & Document_2 : tf("this", d_2) &= \frac{1}{7} \approx 0.14 \\ idf("this", D) &= \log\left(\frac{2}{1}\right) = 0 \\ Document_1 : tfidf("this", d_1) &= 0.2 \times 0 = 0 & Document_2 : tfidf("this", d_2) &= \\ 0.14 \times 0 &= 0 \end{aligned}$$

صفر شدن مقدار $tf-idf$ برای "this" نشان می‌دهد که این کلمه از آنجا که در تمام سندها تکرار شده است بنابراین از اهمیت کمی برخوردار است.

۳-۳-۲ مدل فهرست‌سازی معنایی نهفته

علی‌رغم ویژگی‌های مناسب مدل کاهش‌دهنده‌ی $tf-idf$ مانند شناسایی مجموعه کلمه‌هایی که برای هر سند آن را از دیگر سندها در یک مجموعه متمایز می‌کند، مشکلاتی مانند میزان کاهش ناچیز طول هر سند و همچنین در نظر نگرفتن خصوصیت آماری در داخل هر سند از نقاط ضعف این روش است. برای غلبه بر این مشکلات محققین حوزه‌ی IR چندین مدل کاهش بعد دیگر معرفی کردند که مهمترین آن‌ها مدل فهرست کردن معنایی نهفته^۱ (LSI) است که توسط Deerwester [۲۴] و همکاران در سال ۱۹۹۰ ارائه گردید.

مدل LSI با استفاده از تجزیه مقدار منفرد^۲ بر روی ماتریس خروجی از مدل $tf-idf$ یک زیرفضای خطی^۳ در فضای ویژگی‌های مدل $tf-idf$ شناسایی می‌کند. این روش منجر به کاهش و فشرده‌سازی قابل توجهی در

Term	Term Count
this	1
is	1
another	2
example	3

(ب) سند دوم

Term	Term Count
this	1
is	1
a	2
sample	1

(آ) سند اول

شکل ۳-۳: جدول‌های تعداد تکرار کلمات برای یک مجموعه‌ی اسناد ۲ سندی

¹Latent Semantic Indexing

²Singular Value Decomposition

³Linear Subspace

مجموعه‌های بزرگ می‌گردد. همچنین Deerwester و همکاران ادعا کردند که ویژگی‌های بدست آمده توسط مدل LSI که در حقیقت یک ترکیب خطی از ویژگی‌های مدل $tf-idf$ می‌باشند، توانایی تشخیص بعضی از ویژگی‌های زبانی مانند مترادف و متضاد را دارند [۲۴].

۳-۳-۳ مدل فهرست‌سازی معنایی نهفته‌ی احتمالاتی

برای اثبات ادعاهای مطرح شده توسط مدل LSI (بخش ۳-۳-۲) و همچنین بررسی نقاط ضعف و قدرت این مدل، روش فهرست‌سازی معنایی نهفته‌ی احتمالاتی^۱ (pLSI) توسط Hofmann [۲۵] در سال ۱۹۹۹ معرفی شد. مدل pLSI یک مدل مولد احتمالاتی است که از آن به عنوان یک مدل موضوعی نیز یاد می‌شود [۱]. در روش pLSI هر کلمه‌ی داخل هر سند به عنوان یک نمونه از یک مدل مخلوط مدل می‌شود [۲۵]. مؤلفه‌های این مدل مخلوط در واقع متغیرهای تصادفی چندجمله‌ای می‌باشند که می‌توان آن‌ها را به عنوان یک نمایش از موضوع‌های موجود در سند دانست. بنابراین هر کلمه از یک موضوع خاص تولید می‌شود و کلمه‌های مختلف در داخل یک سند ممکن است از موضوع‌های مختلفی تولید شوند. در نتیجه هر سند به صورت لیستی از این توزیع‌های مخلوط نمایش داده می‌شود و در واقع هر سند به یک مجموعه‌ی از پیش تعیین شده از نظر تعداد از توزیع‌های احتمالاتی کاهش پیدا می‌کند.

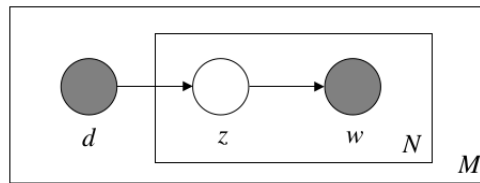
در شکل ۴-۳ مدل pLSI را مشاهده می‌کنید. در شکل ۴-۳ مجموعه سند دارای M سند مختلف است که هر کدام از آن‌ها دارای N کلمه هستند. همچنین در این شکل (۴-۳)، w نشان دهنده‌ی کلمه، و z نماد یک توزیع چندجمله‌ای از موضوع‌ها در یک سند مشخص است. با توجه به شکل ۴-۳ و توضیحات بیان شده، در مدل pLSI فرض بر آن است که یک سند مانند d و یک کلمه مانند w_n در صورت داشتن یک موضوع مشاهده شده مانند z همان‌طور که در رابطه‌ی

$$p(d, w_n) = p(d) \sum_z p(w_n|z)p(z|d) \quad (۲-۳)$$

مشاهده می‌شود دارای استقلال شرطی از یکدیگر هستند.

همان‌طور که پیش از این در بخش ۱۳-۲ بیان شد، یک فرض اساسی که مدل‌های موضوعی بر اساس آن ساخته می‌شوند در نظر گرفتن چند موضوع برای هر سند است. مدل pLSI به عنوان یک مدل موضوعی این امکان را در رابطه‌ی ۲-۳ برای یک سند مشخص مانند d با ترم $p(z|d)$ در نظر می‌گیرد. توجه شود که در اینجا برچسب d یک برچسب ساختگی برای سندهای مجموعه داده‌ی آموزش است، در حقیقت d یک متغیر تصادفی چندجمله‌ای است که مقادیر ممکن برای آن برابر با تعداد سندهای موجود در مجموعه داده‌ی آموزش که مدل

^۱probabilistic Latent Semantic Indexing



شکل ۳-۴: مدل فهرست‌سازی معنایی نهفته‌ی احتمالاتی [۱]

در آن $p(z|d)$ را یاد می‌گیرد، است. به همین دلیل مدل pLSI یک مدل مولد مناسب برای اسناد ناست، چرا که هیچ راهی برای تقابل و تخصیص احتمال به موضوع‌هایی که آن‌ها را در داده‌های آموزش مشاهده نکرده است در این مدل وجود ندارد [۱].

مشکل دیگر مدل pLSI بدلیل استفاده از توزیع‌های احتمالی موجود در داده‌های آموزش است که باعث می‌شود پارامترهایی که باید توسط مدل تخمین زده شوند به صورت خطی با اندازه داده‌های آموزش رشد پیدا کنند [۱].

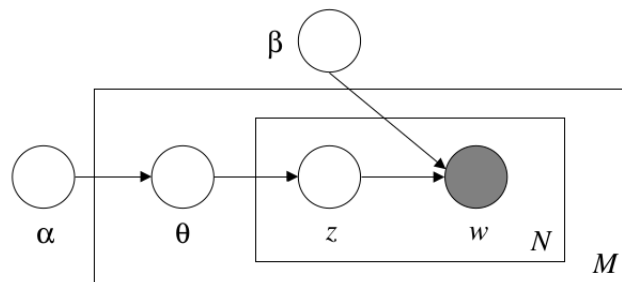
۳-۳-۴ مدل تخصیص دیریکله‌ی پنهان

مدل تخصیص دیریکله‌ی پنهان^۱ (LDA) یک مدل احتمالاتی مولد گرافی است که در سال ۲۰۰۳ توسط Blei و همکاران [۱] معرفی شد و پس از آن به عنوان پایه و اساس مدل‌سازی موضوع در بخش روش‌های گرافی قرار گرفت. مدل LDA را می‌توان معروف‌ترین و مهم‌ترین روش مدل‌سازی موضوع در بخش مدل‌های گرافی و مدل‌های بیزی دانست. تا به امروز روش‌های بسیاری از این مدل برای کاربردهای مختلف مشتق شده‌اند که در بخش‌های ۳-۴-۲ و ۳-۴-۱ دو مورد از آن‌ها را که به تشخیص همزمان موضوع و احساس از داده‌های متنی می‌پردازند را بررسی می‌کنیم.

در روش LDA مانند دیگر روش‌های مدل‌سازی موضوع، هر سند متنی به صورت یک توزیع مخلوط بر روی موضوع‌های مختلف که در آن هر موضوع به وسیله‌ی یک توزیع بر روی کلمه‌ها مشخص می‌شود در نظر گرفته می‌شود [۱]. شکل ۳-۵ مدل LDA استاندارد را نشان می‌دهد.

همان‌طور که در بخش ۳-۳-۳ بیان شد مدل pLSI از مشکلاتی مانند بزرگ شدن فضای پارامترهای مساله به صورت خطی با اندازه داده‌های آموزش و همچنین عدم توانایی در تخصیص احتمال به موضوع‌ها و سندهای از پیش دیده نشده رنج می‌برد. مدل LDA با در نظر گرفتن وزن‌های توزیع مخلوط موضوع‌ها به عنوان یک متغیر تصادفی پنهان، به جای یک مجموعه بزرگ از پارامترهای منحصر به فرد که مستقیماً به داده‌های آموزش متصل باشند بر این دو مشکل غلبه کرده است.

^۱ Latent Dirichlet Allocation



شکل ۳-۵: مدل تخصیص دیریکله‌ی پنهان [۱]

در مدل LDA که در شکل ۳-۵ نشان داده شده است، β یک ماتریس دو بعدی احتمالاتی با اندازه $k \times V$ است که در آن k برابر با تعداد موضوع‌ها و V برابر با اندازه لغت‌نامه یا تعداد کلمات متمایز در متن است. درایه‌های این ماتریس، احتمال حضور هر کلمه در هر موضوع را نشان می‌دهند. برای مثال β_{ij} نشان دهنده‌ی احتمال حضور کلمه‌ی j ام در موضوع i ام است. همان‌طور که در بخش ۲-۴ در مورد مدل‌های گرافی توضیح داده شده در شکل ۳-۵ پارامترهای z ، α ، θ ، β متغیرهای پنهان هستند و w تنها متغیر قابل مشاهده در مدل LDA می‌باشد.

در مدل LDA پارامتر α یک بردار k بعدی که k تعداد موضوع‌ها را نشان می‌دهد، است. α احتمال اولیه‌ی یک توزیع دیریکله است که با نمونه گرفتن از آن با استفاده از رابطه‌ی

$$p(\theta|\alpha) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \theta_1^{\alpha_1-1} \dots \theta_k^{\alpha_k-1} \quad (۳-۳)$$

به ازای هر سند بردار θ که یک توزیع چندجمله‌ای است و نشان دهنده‌ی توزیع موضوع‌ها در هر سند است را بدست می‌آوریم. در رابطه‌ی ۳-۳، Γ تابع گاما است که در بخش ۲-۸ آن را معرفی کردیم.

با داشتن α و β توزیع مشترک پارامترهای θ ، \mathbf{z} و \mathbf{w} از رابطه‌ی

$$p(\theta, \mathbf{z}, \mathbf{w}|\alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^N p(z_n|\theta) p(w_n|z_n, \beta) \quad (۴-۳)$$

بدست می‌آید. در رابطه‌ی ۳-۴، θ یک توزیع چندجمله‌ای موضوعی، بردار \mathbf{z} یک مجموعه N تایی از موضوع‌ها که از θ نمونه گرفته می‌شود و \mathbf{w} یک مجموعه N تایی از کلمه‌ها که مشروط به β و \mathbf{z} نمونه گرفته می‌شود. در ادامه با انتگرال‌گیری روی θ و سیگما روی \mathbf{z} احتمال حاشیه‌ای برای یک سند به فرم رابطه‌ی

$$p(\mathbf{w}|\alpha, \beta) = \int p(\theta|\alpha) \left(\prod_{n=1}^N \sum_{z_n} p(z_n|\theta) p(w_n|z_n, \beta) \right) d\theta \quad (۵-۳)$$

بدست می‌آید. و در نهایت با ضرب احتمال‌های حاشیه‌ای برای هر سند احتمال یک مجموعه سند به شکل

"Arts"	"Budgets"	"Children"	"Education"
NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

The William Randolph Hearst Foundation will give \$1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. "Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services," Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center's share will be \$200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive \$400,000 each. The Juilliard School, where music and the performing arts are taught, will get \$250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual \$100,000 donation, too.

شکل ۳-۶: نمونه خروجی مدل LDA برای چهار موضوع مختلف [۱]

رابطه‌ی

$$p(D|\alpha, \beta) = \prod_{d=1}^M \int p(\theta|\alpha) \left(\prod_{n=1}^N \sum_{z_n} p(z_n|\theta) p(w_n|z_n, \beta) \right) d\theta_d \quad (۳-۶)$$

بدست می‌آید.

برای استنتاج در مدل LDA نمی‌توان از روش‌های مستقیم استفاده کرد و به جای آن باید از روش‌های تقریبی مانند، تقریب لاپلاس^۱، تقریب تغییراتی^۲ و یا روش MCMC که در بخش ۲-۹ معرفی شد استفاده کرد. در مدل معرفی شده توسط آقای Blei همکاران [۱] در سال ۲۰۰۳ از روش تقریب تغییراتی برای استنتاج در مدل LDA استفاده شده است. شکل ۳-۶ خروجی مدل LDA برای چهار موضوع مختلف به همراه کلماتشان در یک متن نمونه را نشان می‌دهد.

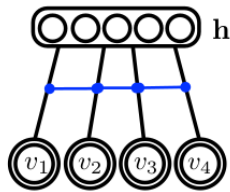
۳-۳-۵ مدل Softmax تکرار شده

مدل Softmax تکرار شده^۳ (RS) که در سال ۲۰۰۹ توسط Hinton و Salakhutdinov [۳] معرفی شد، اولین روش مدل‌سازی موضوع بر پایه‌ی شبکه‌های عصبی است. RS گسترش یافته‌ی مدل RBM است که از آن برای تشخیص توزیع موضوع‌های مختلف در داده‌های متنی استفاده می‌شود. مدل RBM به دلیل محدودیت‌هایی مانند محدود بودن به بردار ورودی باینری و در نظر گرفتن طول ثابت برای ورودی‌ها نمی‌تواند در تشخیص توزیع

¹Laplace Approximation

²Variational Approximation

³Replicated Softmax



شکل ۷-۳: مدل Softmax تکرار شده [۳]

موضوع مورد استفاده قرار بگیرد، چرا که اولاً کلمات باینری نیستند و دوماً در یک مجموعه از داده‌های متنی طول اسناد با یکدیگر متفاوت هستند [۳].

در شکل ۷-۳ مدل RS نشان داده شده است. همان‌طور که در شکل ۷-۳ مشاهده می‌شود مدل RS مانند مدل RBM دارای یک ساختار دو بخشی متشکل از یک لایه قابل مشاهده و یک لایه پنهان است. در مدل RS نیز مانند مدل RBM یک تابع انرژی وابسته به سند ورودی و بردار پنهان آن به شکل رابطه‌ی

$$E(\mathbf{V}, \mathbf{h}) = - \sum_{j=1}^F \sum_{k=1}^K W_j^k h_j \hat{v}^k - \sum_{k=1}^K \hat{v}^k b^k - D \sum_{j=1}^F h_j a_j \quad (۷-۳)$$

تعریف می‌شود که ضمن کمینه کردن آن توزیع موضوع‌های مختلف در متن توسط مدل یاد گرفته می‌شود. احتمال مشاهده‌ی هر سند ورودی در این مدل به کمک رابطه‌ی

$$p(\mathbf{V}) = \frac{1}{Z} \sum_h \exp(-E(\mathbf{V}, \mathbf{h})) \quad (۸-۳)$$

محاسبه می‌گردد که در آن Z همانند مدل RBM تابع قسمت‌بندی است و از رابطه‌ی

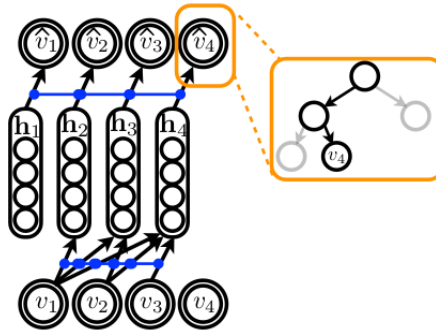
$$Z = \sum_V \sum_h \exp(-E(\mathbf{V}, \mathbf{h})) \quad (۹-۳)$$

بدست می‌آید. در رابطه‌ی ۷-۳، D اندازه سند ورودی، W ماتریس وزن بین لایه‌ی قابل مشاهده و لایه‌ی پنهان، \mathbf{a} و \mathbf{b} به ترتیب بردار بایاس لایه‌های قابل مشاهده و پنهان هستند. همچنین \hat{v}^k که آن را به صورت $\hat{v}^k = \sum_{i=1}^D v_i^k$ تعریف می‌کنیم برابر با تعداد k امین کلمه‌ی لغت‌نامه است. در مدل RS نیز مانند مدل RBM برای آموزش از الگوریتم CD که در بخش ۱۰-۲ معرفی شد استفاده می‌شود.

۶-۳-۳ مدل شبکه‌ی عصبی خود رگرسیو تخمین‌زننده‌ی توزیع سندی

مدل شبکه‌ی عصبی خود رگرسیو تخمین‌زننده‌ی توزیع سندی^۱ (DocNADE) که در شکل ۸-۳ مشاهده می‌شود یک روش بدون نظارت برای مدل‌سازی موضوع بر پایه‌ی شبکه‌های عصبی است. این مدل در سال ۲۰۱۲ توسط Lauly و Larochelle [۴] با ترکیب مدل‌های RS و NADE معرفی شد.

^۱ Document Neural Autoregressive Distribution Estimator



شکل ۳-۸: مدل شبکه‌ی عصبی خود رگرسیو تخمین‌زننده‌ی توزیع سندی [۴]

بردار ورودی در این مدل بر خلاف مدل NADE که در آن بردار ورودی می‌بایست حتماً باینری باشد، یک بردار چندجمله‌ای به شکل بردار ورودی در مدل RS است. اما در این مدل مانند مدل NADE احتمال هر کلمه در داخل سند به شرط تمام کلمات قبل از آن بدست می‌آید [۴]. تفاوت دیگر مدل DocNADE با مدل NADE در نحوه‌ی بدست آوردن احتمال مشاهده‌ی هر کلمه به شرط کلمات قبل است. در مدل DocNADE احتمال در لایه‌ی نهایی با استفاده از یک ساختار درختی محاسبه می‌گردد. در درخت مورد نظر که در شکل ۳-۸ مشخص شده تمام کلمات در برگ‌های آن قرار می‌گیرند و هر مسیر از ریشه تا برگ برابر با احتمال مشاهده‌ی کلمه‌ی متناظر با آن است. استفاده از یک چنین ساختار درختی موجب می‌گردد تا بدست آوردن احتمال هر کلمه نسبت به اندازه لغت‌نامه مقداری لگاریتمی داشته باشد که در کاربردهای عملی بسیار کارآمدتر از حالت مستقیم که خطی است است [۴].

رابطه‌های

$$\mathbf{h}_i(\mathbf{v}_{<i}) = \text{sigm}(\mathbf{c} + \sum_{k<i} \mathbf{w}_{:,v_k}) \quad (۱۰-۳)$$

و

$$p(v_i = w | \mathbf{v}_{<i}) = \frac{\exp(b_w + \mathbf{V}_{w,:} \cdot \mathbf{h}_i(\mathbf{v}_{<i}))}{\sum_{w'} \exp(b_{w'} + \mathbf{V}_{w',:} \cdot \mathbf{h}_i(\mathbf{v}_{<i}))} \quad (۱۱-۳)$$

به ترتیب نحوه‌ی به دست آوردن مقادیر لایه‌ی پنهان و احتمال مشاهده‌ی هر کلمه به شرط تمام کلمات پیشین در مدل DocNADE را نشان می‌دهند. در جدول ۳-۴ نمونه‌ای از چهار موضوع یاد گرفته شده توسط مدل DocNADE به همراه ده کلمه‌ای که در هر موضوع بیشترین احتمال را داشته‌اند مشاهده می‌شود.

۳-۴ مدل‌های مشترک موضوع و احساس

تمام مدل‌های بررسی شده در بخش‌های پیشین تنها توانایی تشخیص موضوع از داده‌های متنی را داشتند. گروه دیگری از مدل‌های موضوعی وجود دارند که به صورت همزمان به تشخیص موضوع‌ها و احساس همراه با هر کدام

می‌پردازند. در بخش های ۳-۴-۲ و ۳-۴-۱ دو مدل موجود در این زمینه، که یکی به تشخیص موضوع و احساس آن در سطح کلمه (۳-۴-۲) و یکی در سطح جمله (۳-۴-۱) می‌پردازد را بررسی می‌کنیم. همچنین در بخش ۳-۴-۳ یک مدل با رویکردی نظارت شده که به تازگی برای تشخیص موضوع و احساس در داده‌های متنی پیشنهاد شده است را معرفی می‌کنیم.

۳-۴-۱ مدل یکی سازی احساس-موضوع

پیدا کردن جنبه‌هایی از یک محصول که کاربران در بازبینی‌های آنلاین خود مورد ارزیابی قرار می‌دهند همواره کار مشکلی بوده است. علاوه بر تشخیص موضوع‌ها و جنبه‌های موجود در یک بازبینی آنلاین، تشخیص احساس همراه با هر کدام نیز برای شرکت سازنده و افراد دیگری که به دنبال کسب اطلاعات در مورد یک محصول خاص هستند، اهمیت دارد. مدل یکی سازی احساس-موضوع^۱ (ASUM) در سال ۲۰۱۱ برای تشخیص موضوع‌ها و احساس همراه با آن‌ها در بازبینی‌های آنلاین توسط Jo و Oh معرفی شد [۵].

شکل ۳-۹ مدل ASUM را نشان می‌دهد. این مدل همان‌طور که در شکل ۳-۹ مشاهده می‌گردد گسترش یافته‌ی مدل LDA است [۵] و در گروه مدل‌های احتمالاتی گرافی مولد قرار می‌گیرد. در مدل LDA هر کلمه به صورت مجزا از یک موضوع نمونه گرفته می‌شود و فرض بر آن است که هر کلمه می‌تواند موضوع خود را داشته باشد، اما در مدل ASUM فرض بر آن است که هر جمله دارای یک موضوع است و تمام کلمات یک جمله از یک موضوع نمونه گرفته می‌شوند [۵].

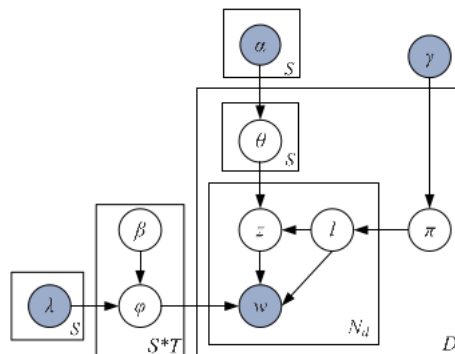
در مدل ASUM ما برای هر سند یک توزیع چندجمله‌ای احساسی و برای هر یک از احساس‌ها یک توزیع چندجمله‌ای موضوعی داریم. در حالت مولد پس از نمونه گرفتن از توزیع احساسی و توزیع موضوعی متناسب با احساس انتخاب شده برای سند جاری، برای هر جمله یک احساس و یک موضوع نمونه گرفته می‌شود و سپس

چهار موضوع یادگرفته شده توسط DocNADE			
christianity	team	shuttle	card
god	games	orbit	driver
p&g	seasons	lunar	drivers
jesus	baseball	spacecraft	bus
bible	players	nasa	video
faith	game	space	vga
muslim	hockey	launch	monitor
christ	play	saturn	ibm
atheist	teams	billion	cards
christians	sale	satellite	ram

جدول ۳-۴: نمونه‌ای از موضوع‌های یادگرفته شده توسط مدل DocNADE [۴]

^۱ Aspect and Sentiment Unification Model

شده و مشروط به آن یک موضوع از توزیع موضوعی $\theta_{d,l}$ نمونه گرفته می‌شود. در نهایت با داشتن برچسب موضوع احساس، از توزیع احتمالاتی مشروط به موضوع و احساس برای هر مجموعه سند یک کلمه تولید می‌شود.



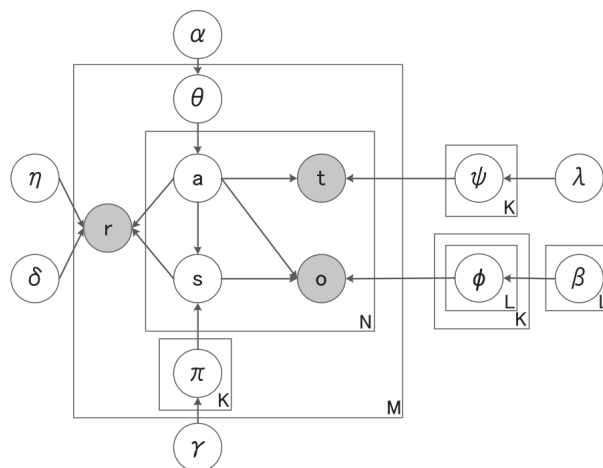
شکل ۳-۱۰: مدل نظارت‌شده‌ی ضعیف تشخیص مشترک احساس/موضوع [۶]

لازم به ذکر است که در مدل LDA به ازای هر سند یک توزیع احتمالاتی برای موضوع‌ها وجود دارد، اما در مدل JST به ازای هر برچسب احساس یک توزیع احتمالاتی برای موضوع‌ها در سند تعریف می‌شود. به طور مثال اگر سه برچسب احساس مثبت، منفی یا بی‌طرف در نظر گرفته شود، برای هر سند سه توزیع احتمالاتی وجود خواهد داشت که هر کدام متناظر با یک احساس خواهند بود. تفاوت دیگری که بین مدل LDA و JST است از این قرار است که در LDA به هنگام نمونه‌برداری از کلمه، این کار تنها مشروط به توزیع موضوعی برای مجموعه سند انجام می‌شود ولی در JST این نمونه‌برداری مشروط به هر دو توزیع موضوعی و احساسی انجام می‌شود [۶]. در شکل ۳-۱۱ می‌توان خروجی مدل JST برای دو احساس مثبت ۳-۱۱آ و منفی ۳-۱۱ب را به همراه دو موضوع برای هر احساس و همچنین پانزده کلمه‌ای که بیشترین احتمال در هر موضوع داشتند را مشاهده می‌کنید.

prison	bad	ship	good
evil	worst	titan	realli
guard	plot	crew	plai
green	stupid	cameron	great
hank	act	alien	just
wonder	suppos	jack	perform
excute	script	water	nice
secret	wast	stori	fun
mile	dialogu	rise	lot
death	bore	rose	act
base	poor	boat	direct
tom	complet	deep	best
convict	line	ocean	get
return	terribl	dicaprio	entertain
franklin	mess	sink	better

(آ) نمونه خروجی مدل JST برای احساس مثبت (ب) نمونه خروجی مدل JST برای احساس منفی

شکل ۳-۱۱: خروجی مدل JST برای دو احساس مثبت و منفی و دو موضوع برای هر احساس و پانزده کلمه در هر موضوع [۶]



شکل ۳-۱۲: مدل نظارت‌شده‌ی مشترک موضوع و احساس [۷]

۳-۴-۳ مدل نظارت‌شده‌ی مشترک موضوع و احساس

مدل نظارت‌شده‌ی مشترک موضوع و احساس (SJASM)^۱ که در شکل ۳-۱۲ نشان داده شده است در سال ۲۰۱۷ توسط Hai و همکاران معرفی شد [۷]. SJASM یک رویکرد مولد احتمالی است که مانند مدل‌های ASUM و JST بر پایه‌ی روش LDA است و در دسته‌ی روش‌های بیزی قرار می‌گیرد. SJASM با اضافه کردن چندین لایه و تغییر در ساختار مدل LDA از برای تشخیص همزمان موضوع و احساس در داده‌های متنی مورد استفاده قرار می‌گیرد [۷].

در مدل SJASM هر سند تولید شده توسط کاربران به صورت جفت کلمه‌های موضوع و احساس نمایش داده می‌شود. SJASM با استفاده از این جفت کلمه‌ها به مدل‌سازی مشترک موضوع و احساس در داده‌های متنی می‌پردازد [۷].

۳-۵ مدل‌های چندحالت

مدل‌های چندحالت^۲ دسته‌ای از مدل‌های موضوعی هستند که داده‌ی ورودی در آن‌ها ترکیبی از چند حالت مختلف داده است. در تمام مدل‌های معرفی شده تا این قسمت از پژوهش، داده‌ی ورودی تنها اسناد متنی بودند یا به عبارت دیگر تنها یک حالت از داده به عنوان ورودی به مدل وارد می‌شود، اما مدل‌های چندحالت مانند آنچه که در بخش ۳-۵-۱ معرفی می‌شود بر روی ترکیب همزمان دو یا چند حالت مختلف از داده مثلاً تصویر و متن همراه با آن عمل می‌کنند.

^۱Supervised Joint Aspect and Sentiment Model

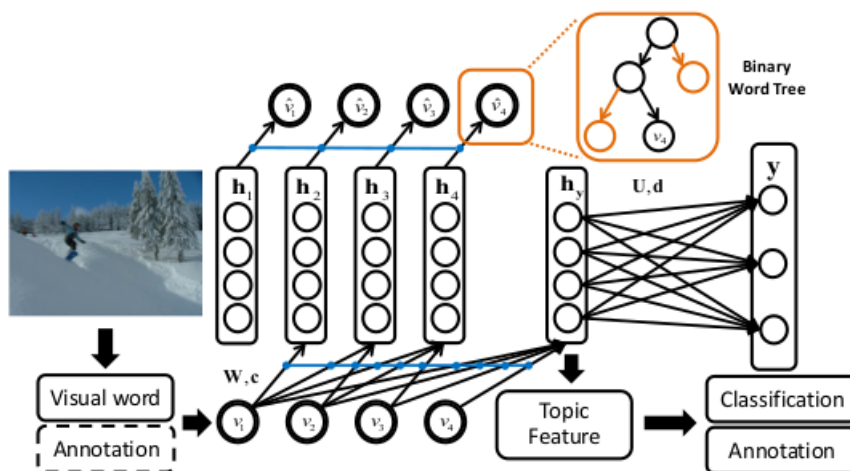
^۲Multimodal

۳-۵-۱ مدل نظارت‌شده‌ی شبکه‌ی عصبی خود رگرسیو تخمین‌زننده‌ی توزیع سندی

مدل نظارت‌شده‌ی شبکه‌ی عصبی خود رگرسیو تخمین‌زننده‌ی توزیع سندی^۱ (SupDocNADE) که در شکل ۳-۱۳ مشاهده می‌شود، در سال ۲۰۱۴ توسط Zheng و همکاران [۸] معرفی شد. این مدل گسترش یافته‌ی مدل DocNADE می‌باشد که در بخش ۳-۳-۶ معرفی شد [۸]. در مدل SupDocNADE داده‌های ورودی تنها اسناد متنی نیستند. ورودی در این مدل تصاویر به همراه توضیح کوتاهی در مورد هر تصویر است که مدل ترکیب این دو نوع داده در کنار یکدیگر را یاد می‌گیرد. در این مدل ابتدا هر تصویر با استفاده الگوریتم‌های موجود در حوزه‌ی پردازش تصویر تغییر یافته و به یک سند متنی تبدیل شده و به توضیح مربوط به آن متصل می‌شود و به عنوان بردار ورودی به مدل وارد می‌شود [۸]. تفاوت دیگر این مدل با مدل DocNADE اضافه کردن یک ساختار بر بایه‌ی شبکه‌ی عصبی در کنار قسمت‌های موجود است که با استفاده از ویژگی‌های یاد گرفته شده از ترکیب هر تصویر و متن مربوط به آن عمل طبقه‌بندی را انجام می‌دهد [۸].

۳-۶ نتیجه‌گیری

در این بخش پژوهش‌های پیشین در زمینه‌ی تخمین توزیع، مدل‌سازی موضوع، مدل‌سازی احساس-موضوع به صورت مشترک و همچنین مدل‌سازی داده‌های چندوجهی مورد بررسی قرار گرفت. ساختارهای بررسی شده در این بخش در دو دسته‌ی مدل‌های شبکه‌های عصبی و مدل‌های گرافی بیزین قرار می‌گیرند. در این پژوهش با ایده گرفتن از ساختارهای گرافی بیزین و استفاده از مدل‌های شبکه‌های عصبی هدف ساخت مدلی برای شبیه‌سازی مشترک احساس و موضوع در داده‌های متنی است. در بخش بعدی کلیات نظری مدل پیشنهادی در این پایان‌نامه



شکل ۳-۱۳: مدل نظارت‌شده‌ی شبکه‌ی عصبی خود رگرسیو تخمین‌زننده‌ی توزیع سندی [۸]

^۱Supervised Document Neural Autoregressive Distribution Estimator

و روابط و ساختار آن به صورت دقیق معرفی و بررسی می شوند.

فصل چهارم

مدل پیشنهادی

۴-۱ مقدمه

در این بخش روش پیشنهادی در این پژوهش که یک مدل بر پایه‌ی شبکه‌های عصبی برای تشخیص همزمان احساس و موضوع از داده‌های متنی است ارائه می‌گردد. همانطور که پیش از این نیز ذکر شد مدل پیشنهادی در این پژوهش گسترش یافته و تلفیقی از چند مدل شناخته شده در زمینه‌ی مدل‌سازی احساس و موضوع و همچنین تشخیص توزیع‌های احتمالی موجود در داده‌های ورودی است، لذا در بخش‌های پیش‌رو ضمن معرفی کامل این مدل‌ها و تعریف ساختار و نحوه‌ی عملکرد هرکدام به مدل پیشنهادی در این پایان‌نامه می‌رسیم و آن را به طور کامل مورد بررسی قرار داده و تعریف می‌کنیم.

۴-۲ مدل پایه

مدل پایه در این پژوهش شبکه عصبی ماشین بلتزن محدود است که پیش از این در بخش ۳-۲-۱ آن را RBM نامیدیم و به صورت مختصر معرفی گردید. RBM که در شکل ۴-۱ نشان داده شده است، یک مدل بدون نظارت برای داده‌های باینری است که در دسته‌ی مدل‌های مولد احتمالاتی قرار می‌گیرد. در این مدل با

بیشینه کردن یک تابع انرژی، یا کمینه کردن مقدار منفی آن که به صورت رابطه‌ی

$$E(\mathbf{v}, \mathbf{h}) = - \sum_i \sum_j v_i W_{ij} h_j - \sum_i v_i a_i - \sum_j h_j b_j \quad (۱-۴)$$

تعریف می‌شود، توزیع‌های احتمالی موجود در داده‌های ورودی یاد گرفته می‌شود و از داده‌های ورودی ویژگی استخراج می‌گردد. در رابطه‌ی ۱-۴ مجموعه‌ی پارامترهای مدل به صورت $\theta = \{W, \mathbf{a}, \mathbf{b}\}$ است. $W_{D \times H}$ ماتریس وزن بین لایه‌ی ورودی و لایه‌ی پنهان است، که در آن D اندازه بردار ورودی و H اندازه لایه‌ی پنهان هستند. \mathbf{a} بردار بایاس لایه‌ی ورودی با اندازه D و \mathbf{b} بردار بایاس لایه‌ی پنهان با اندازه H است. همان‌طور که در بخش ۱-۲-۳ نیز بیان گردید در مدل RBM داده‌های ورودی و لایه‌ی پنهان متناظر با آن که توسط مدل از بردار ورودی بدست می‌آید هر دو در حالت باینری (۰ یا ۱) هستند که این امر یکی از محدودیت‌های این روش است.

در مدل RBM احتمال هر ترکیب (\mathbf{v}, \mathbf{h}) از رابطه‌ی

$$p(\mathbf{v}, \mathbf{h}) = \frac{1}{Z(\theta)} e^{-E(\mathbf{v}, \mathbf{h})} \quad (۲-۴)$$

بدست می‌آید که در آن $Z(\theta)$ تابع قسمت‌بندی است که مقدار آن با استفاده از رابطه‌ی

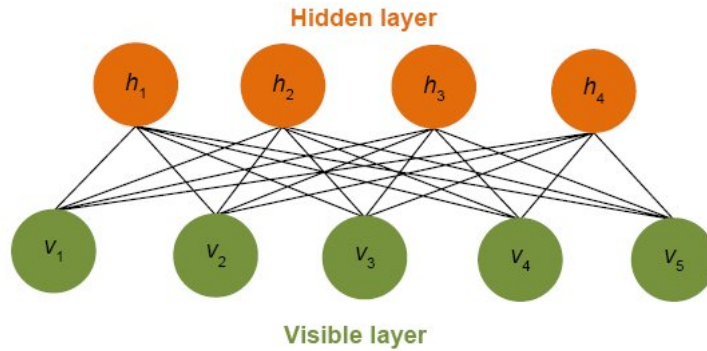
$$Z = \sum_{\mathbf{v}, \mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h})} \quad (۳-۴)$$

محاسبه می‌شود و تضمین می‌کند که مقدار بدست آمده برای هر ترکیب (\mathbf{v}, \mathbf{h}) یک مقدار صحیح احتمالی (بین ۰ و ۱) است. در این مدل احتمال هر بردار ورودی از رابطه‌ی

$$p(\mathbf{v}) = \sum_{\mathbf{h}} \frac{1}{Z(\theta)} e^{-E(\mathbf{v}, \mathbf{h})} \quad (۴-۴)$$

بدست می‌آید.

برای آموزش این مدل از الگوریتم CD که در بخش ۱۰-۲ معرفی گردید استفاده می‌شود. در مراحل آموزش و همچنین آزمودن این مدل نیاز به محاسبه‌ی مقادیر لایه‌ی پنهان مشروط به بردار ورودی و برعکس است. برای



شکل ۴-۱: ماشین بلترزن محدود

محاسبه $p(\mathbf{h}|\mathbf{v})$ داریم:

$$\begin{aligned}
 p(\mathbf{h}|\mathbf{v}) &= \frac{p(\mathbf{h}, \mathbf{v})}{p(\mathbf{v})} = \frac{\frac{1}{Z} e^{-E(\mathbf{v}, \mathbf{h})}}{\sum_{\mathbf{h}} p(\mathbf{v}, \mathbf{h})} \\
 &= \frac{\frac{1}{Z} e^{\mathbf{v}W\mathbf{h}^T + \mathbf{a}^T\mathbf{v} + \mathbf{b}^T\mathbf{h}}}{\sum_{\mathbf{h}} \frac{1}{Z} e^{\mathbf{v}W\mathbf{h}^T + \mathbf{a}^T\mathbf{v} + \mathbf{b}^T\mathbf{h}}} \\
 &= \frac{e^{\mathbf{v}W\mathbf{h}^T} \cdot e^{\mathbf{a}^T\mathbf{v}} \cdot e^{\mathbf{b}^T\mathbf{h}}}{\sum_{\mathbf{h}} e^{\mathbf{v}W\mathbf{h}^T} \cdot e^{\mathbf{a}^T\mathbf{v}} \cdot e^{\mathbf{b}^T\mathbf{h}}} = \frac{e^{\mathbf{v}W\mathbf{h}^T + \mathbf{b}^T\mathbf{h}}}{\sum_{\mathbf{h}} e^{\mathbf{v}W\mathbf{h}^T + \mathbf{b}^T\mathbf{h}}} = \frac{e^{\mathbf{v}W\mathbf{h}^T + \mathbf{b}^T\mathbf{h}}}{Z'} \\
 &= \frac{1}{Z'} \exp \{ \mathbf{v}W\mathbf{h}^T + \mathbf{b}^T\mathbf{h} \} = \frac{1}{Z'} \exp \left\{ \sum_{j=1}^H \mathbf{v}W_j h_j + \sum_{j=1}^H b_j h_j \right\} \\
 \Rightarrow p(\mathbf{h}|\mathbf{v}) &= \frac{1}{Z'} \prod_{j=1}^H \exp \{ \mathbf{v}W_j h_j + b_j h_j \}
 \end{aligned} \tag{۵-۴}$$

در ترم نهایی در رابطه‌ی ۵-۴ مشاهده می‌شود که مقدار $p(\mathbf{h}|\mathbf{v})$ برابر است با حاصلضرب \exp به ازای تمام ابعاد بردار \mathbf{h} ، لذا نتیجه گرفته می‌شود که ابعاد بردار \mathbf{h} یعنی همان h_j ها نسبت به یکدیگر دارای استقلال شرطی هستند. در نتیجه می‌توان رابطه‌ی ۵-۴ را به شکل رابطه‌ی

$$p(\mathbf{h}|\mathbf{v}) = \prod_{j=1}^H p(h_j|\mathbf{v}) \tag{۶-۴}$$

بازنویسی کرد. همچنین داریم:

$$\begin{aligned}
 p(h_j = 1|\mathbf{v}) &= \frac{p(h_j = 1, \mathbf{v})}{p(h_j = 1, \mathbf{v}) + p(h_j = 0, \mathbf{v})} = \frac{\exp\{\mathbf{v}W_j + b_j\}}{\exp\{0\} + \exp\{\mathbf{v}W_j + b_j\}} \\
 &= \text{sigmoid}(\mathbf{v}W_j + b_j)
 \end{aligned} \tag{۷-۴}$$

$$p(\mathbf{h}|\mathbf{v}) = \prod_{j=1}^H \text{sigmoid}(\mathbf{v}W_j + b_j) \tag{۸-۴}$$

به همین شکل می‌توان نتیجه گرفت:

$$p(\mathbf{v}|\mathbf{h}) = \prod_{i=1}^D \text{sigmoid}(W_i \mathbf{h} + a_i) \quad (9-4)$$

برای آموزش این مدل فرض می‌شود یک مجموعه داده‌ی باینری به صورت $\{\mathbf{v}^t\}_{t=1}^n$ به عنوان داده‌های آموزش در اختیار است. هدف پیدا کردن پارامترهای $W, \mathbf{a}, \mathbf{b}$ به صورتی است که لگاریتم درست‌نمایی برای مشاهده‌ی داده‌های آموزش بیشینه باشد. به بیان ریاضی می‌خواهیم مقدار رابطیه‌ی

$$\begin{aligned} \ell(W, \mathbf{a}, \mathbf{b}) &= \sum_{t=1}^n \log p(\mathbf{v}^t) = \sum_{t=1}^n \log \sum_h p(\mathbf{v}^t, h) \\ &= \sum_{t=1}^n \log \frac{1}{Z} \sum_h e^{-E(\mathbf{v}^t, h)} = \sum_{t=1}^n \log \sum_h e^{-E(\mathbf{v}^t, h)} - n \log Z \\ &= \sum_{t=1}^n \log \sum_h e^{-E(\mathbf{v}^t, h)} - n \log \sum_{\mathbf{v}, \mathbf{h}} \exp \{-E(\mathbf{v}^t, h)\} \end{aligned} \quad (10-4)$$

را نسبت به مجموعه پارامترهای $\theta = \{W, \mathbf{a}, \mathbf{b}\}$ بیشینه کنیم.

رابطه‌ی ۱۰-۴ را معادله‌ی لگاریتم درست‌نمایی تعریف می‌کنیم که برای بیشینه کردن آن نسبت به مجموعه پارامترهای θ باید از آن نسبت به پارامترهای این مجموعه مشتق گرفت.

$$\begin{aligned} \nabla_{\theta} \ell(\theta) &= \nabla_{\theta} \sum_{t=1}^n \log \sum_h \exp \{-E(\mathbf{v}^t, \mathbf{h})\} - n \nabla_{\theta} \log \sum_{\mathbf{v}, \mathbf{h}} \exp \{-E(\mathbf{v}, \mathbf{h})\} \\ &= \sum_{t=1}^n \frac{\sum_h \exp \{-E(\mathbf{v}^t, \mathbf{h})\} \nabla_{\theta} - E(\mathbf{v}^t, \mathbf{h})}{\sum_h \exp \{-E(\mathbf{v}^t, \mathbf{h})\}} - n \frac{\sum_{\mathbf{v}, \mathbf{h}} \exp \{-E(\mathbf{v}, \mathbf{h})\} \nabla_{\theta} - E(\mathbf{v}, \mathbf{h})}{\sum_{\mathbf{v}, \mathbf{h}} \exp \{-E(\mathbf{v}, \mathbf{h})\}} \\ &= \sum_{t=1}^n E_{p(\mathbf{h}|\mathbf{v}^t)}[\nabla_{\theta} - E(\mathbf{v}^t, \mathbf{h})] - n E_{p(\mathbf{v}, \mathbf{h})}[\nabla_{\theta} - E(\mathbf{v}, \mathbf{h})] \end{aligned} \quad (11-4)$$

$$\nabla_W - E(\mathbf{v}, \mathbf{h}) = \mathbf{v}^T \mathbf{h} \quad \nabla_{\mathbf{a}} - E(\mathbf{v}, \mathbf{h}) = \mathbf{v} \quad \nabla_{\mathbf{b}} - E(\mathbf{v}, \mathbf{h}) = \mathbf{h} \quad (12-4)$$

ترم اول سمت راست رابطه ۱۱-۴ یک مقدار مورد انتظار برای مشتق تابع انرژی نسبت به توزیع شرطی داده‌های آموزش، $p(\mathbf{h}|\mathbf{v})$ ، و ترم دوم آن یک مقدار مورد انتظار برای تابع انرژی نسبت به توزیع مشترک برای تمام حالت‌های مدل $p(\mathbf{v}, \mathbf{h})$ است. با توجه به رابطه‌ی ۱۲-۴ برای بدست آوردن ترم اول در رابطه‌ی

$$\nabla_W \ell(W, \mathbf{a}, \mathbf{b}) = \sum_{t=1}^n \mathbf{v}^t \hat{\mathbf{h}}^t - n E_{p(\mathbf{v}, \mathbf{h})}[\mathbf{v}^T \mathbf{h}] \quad (13-4)$$

برای هر بردار \mathbf{v} مقدار \mathbf{h} متناظر با آن را با استفاده از رابطه‌ی ۸-۴ بدست می‌آوریم. از آن‌جا که ترم دوم رابطه‌ی ۱۱-۴ به توزیع تمام حالات موجود برای مدل $(p(\mathbf{v}, \mathbf{h}))$ وابسته است محاسبه‌ی آن برای ما امکان‌پذیر نیست و

باید مقدار آن را به صورت تقریبی بدست آوریم. الگوریتم CD در اینجا وارد عمل می‌شود و مقدار مورد انتظار از مشتق تابع انرژی بر روی یک توزیع مشترک را با یک تخمین نقطه‌ای جایگزین می‌کند. به این صورت که یک بردار از داده‌های آموزش انتخاب کرده و با استفاده از الگوریتم نمونه‌برداری Gibbs و همچنین روابط ۹-۴ و ۸-۴ مقادیر بردارهای \mathbf{h} و \mathbf{v} را به ترتیب محاسبه می‌کند. یک چرخه‌ی کامل الگوریتم نمونه‌برداری Gibbs شامل محاسبه‌ی بردار \mathbf{h} از یک بردار \mathbf{v} و سپس ساختن مجدد بردار \mathbf{v} از این بردار \mathbf{h} است. این چرخه تا زمانی که به یک حالت تعادل دست یابید ادامه پیدا می‌کند، در حالت تعادل مقدار بردارهای \mathbf{v} و \mathbf{h} بدست آمده برابر با مقدار مورد انتظار برای ترم دوم رابطه ۴-۱۳ در نظر گرفته می‌شود. همچنین با استفاده از مقادیر بدست آمده برای بردارهای \mathbf{v} و \mathbf{h} مقدار روابط

$$\nabla_{\mathbf{a}} \ell(W, \mathbf{a}, \mathbf{b}) = \sum_{t=1}^n \mathbf{v}^t - n E_{p(\mathbf{v}, \mathbf{h})}[\mathbf{v}] \quad (14-4)$$

و

$$\nabla_{\mathbf{b}} \ell(W, \mathbf{a}, \mathbf{b}) = \sum_{t=1}^n \mathbf{h}^t - n E_{p(\mathbf{v}, \mathbf{h})}[\mathbf{h}] \quad (15-4)$$

نیز محاسبه می‌گردد.

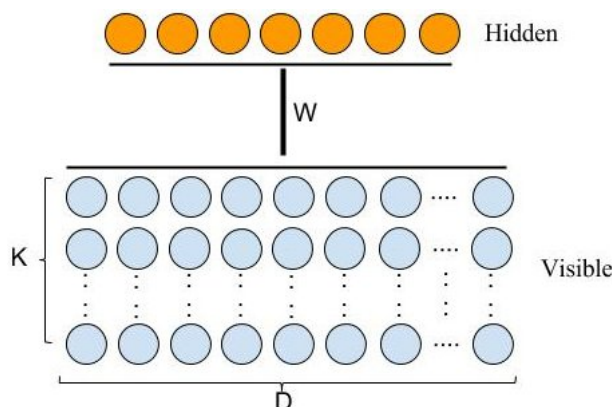
همان‌طور که پیش از این در بخش ۲-۱۰ بیان شد، آقای Hinton نشان داد که تنها تعداد کمی چرخه‌ی نمونه‌برداری Gibbs برای بدست آوردن تقریبی مناسب نسبت به توزیع $p(\mathbf{v}, \mathbf{h})$ کافی است [۱۰][۱۲]. اگرچه پیاده‌سازی‌های انجام شده نشان داده است که تنها انجام یک چرخه‌ی Gibbs برای محاسبه‌ی یک تقریب صحیح از جهت حرکت گرادیان کافی است [۱۲]. در نهایت برای کمینه کردن تابع انرژی رابطه‌ی بروزرسانی کردن مجموعه پارامترهای θ به شکل رابطه‌ی

$$\theta_{t+1} = \theta_t + \epsilon (E_{P_{data}}[\nabla_{\theta} - E(\mathbf{v}, \mathbf{h})] - E_{P_{model}}[\nabla_{\theta} - E(\mathbf{v}, \mathbf{h})]) \quad (16-4)$$

می‌گردد که در آن ϵ ضریب یادگیری است و به صورت تجربی مشخص گردد.

۳-۴ ماشین بلتزن محدود با واحدهای قابل مشاهده صحیح

ساختار معرفی شده در بخش ۲-۴ مدل استاندارد RBM است که به صورت دقیق مورد بررسی قرار گرفت و به عنوان مدل پایه برای روش پیشنهادی در این پژوهش در نظر گرفته می‌شود. برای تعریف ساختار مورد نظر در این پایان‌نامه جهت مدل‌سازی احساس و موضوع نیاز به معرفی و بررسی نمونه‌های پیچیده‌تری از مدل RBM استاندارد است.



شکل ۴-۲: ماشین بلترمن محدود با واحدهای قابل مشاهده صحیح

محدود بودن به حالت باینری برای داده‌های ورودی و همچنین طول ثابت برای آن‌ها دو اشکال اساسی در مدل RBM استاندارد هستند. حال فرض کنید ساختاری داریم که در آن داده‌های ورودی همچنان دارای طول ثابت هستند اما به جای حالت باینری می‌توانند هر مقدار صحیح غیر منفی از یک حداقل تا یک حداکثر را اختیار کنند. در این ساختار که در شکل ۴-۲ نشان داده شده است به جای بردار ورودی ما ماتریس ورودی خواهیم داشت، و هر داده به صورت یک ماتریس با درایه‌های ۰ و ۱ به مدل وارد می‌شود. در مدل RBM استاندارد ورودی تنها یک بردار با طول ثابت با درایه‌های ۰ و ۱ بود، که حضور و یا عدم حضور هر یک از ویژگی‌های متناظر با درایه‌ی مورد نظر در بردار ورودی را مشخص می‌کند. اما در این مدل هر داده‌ی ورودی به صورت ماتریسی متشکل از ۰ یا ۱ است.

فرض کنید در مساله‌ای که با آن سرو کار داریم هر داده‌ی ورودی دارای D ویژگی است که هر یک از این ویژگی‌ها می‌توانند K مقدار داشته باشند. مدل RBM استاندارد توانایی کار کردن با یک چنین داده‌های ورودی را ندارد چرا که در آن داده‌های ورودی تنها می‌توانند یک بردار با طول ثابت و شامل ۰ و ۱ باشند. در این مدل جدید هر داده‌ی ورودی به صورت ماتریسی با اندازه $K \times D$ در نظر گرفته می‌شود که همان‌طور که بیان شد D طول بردار ورودی یا همان تعداد ویژگی‌های مساله و K بیشینه مقداری است که هر ویژگی می‌تواند داشته باشد. برای ساخت این ماتریس ابتدا تمام درایه‌های آن را صفر می‌کنیم و سپس برای هر ستون یا به عبارتی برای هر ویژگی سطر متناظر با مقدار آن ویژگی را ۱ می‌کنیم. در این حالت به ازای هر داده‌ی ورودی یک ماتریس شامل ۰ و ۱ داریم با این خصوصیت که در هر ستون آن تنها مقدار یک سطر ۱ است و مابقی سطرها مقدارشان ۰ است.

تا به اینجا تفاوت این ساختار با مدل RBM استاندارد را معرفی کردیم. از نظر تفاوت در فضای پارامتری، در مدل RBM استاندارد ماتریس وزن بین لایه‌ی قابل مشاهده و لایه‌ی پنهان یک ماتریس دو بعدی بود، اما در

این ساختار به دلیل تغییر در شکل ورودی ماتریس وزن بین این دو لایه یک ماتریس سه بعدی است. در این ساختار نیز مانند مدل RBM استاندارد که در آن تمام ابعاد بردار ورودی به تمام واحدهای لایه پنهان متصل بودند، در اینجا نیز یک اتصال کامل بین تمام قسمت‌های ماتریس ورودی و تمام واحدهای لایه پنهان برقرار است. تفاوت دیگر این دو ساختار در فضای پارامترها مربوط به بایاس لایه قابل مشاهده است. برخلاف مدل RBM استاندارد که برای لایه قابل مشاهده یک بردار بایاس وجود داشت در ساختار نشان داده شده در شکل ۲-۴ ما یک ماتریس دو بعدی برای بایاس خواهیم داشت که در فرآیند آموزش درایه‌های آن یاد گرفته می‌شوند. با توجه به تغییرات بیان شده در ساختار RBM استاندارد و همچنین تغییرات اعمال شده در فضای پارامترهای مساله، روابط ۱-۴ تا ۹-۴ نیز دچار تغییرات اساسی می‌گردند که در اینجا به بیان آن‌ها می‌پردازیم. فرض کنید قصد مدل‌سازی بردارهای ورودی \mathbf{v} با مدلی مانند آنچه در شکل ۲-۴ نشان داده شده است را داریم. اگر هر داده‌ی \mathbf{v} به شکل $\mathbf{v} \in \{1, \dots, K\}^D$ باشد، که در آن همان‌طور که پیش از این بیان کردیم، K بیشینه مقداری است که هر ویژگی می‌تواند دریافت کند و D اندازه داده‌های ورودی است. همچنین بردار پنهان به شکل $\mathbf{h} \in \{0, 1\}^H$ می‌باشد. اگر \mathbf{V} ماتریس باینری قابل مشاهده باشد که برای هر داده‌ی ورودی ساخته می‌شود، در این صورت تابع انرژی برای حالت $\{\mathbf{V}, \mathbf{h}\}$ به شکل

$$E(\mathbf{V}, \mathbf{h}) = - \sum_{i=1}^D \sum_{j=1}^H \sum_{k=1}^K W_{ijk} h_j v_{ik} - \sum_{i=1}^D \sum_{k=1}^K v_{ik} a_{ik} - \sum_{j=1}^H h_j b_j \quad (۱۷-۴)$$

تعریف می‌شود.

در رابطه‌ی ۱۷-۴ نیز مانند حالت قبل $\{W, a, \mathbf{b}\}$ مجموعه پارامترهای مدل هستند که در آن $W_{D \times H \times K}$ ماتریس سه بعدی وزن بین واحد قابل مشاهده‌ی i ام و مقدار k ام و واحد j ام بردار لایه پنهان است، همچنین در رابطه‌ی ۱۷-۴، $a_{D \times K}$ ماتریس دو بعدی بایاس برای لایه قابل مشاهده و $\mathbf{b}_{1 \times H}$ بردار بایاس برای لایه پنهان است. احتمالی که این ساختار به ماتریس باینری قابل مشاهده‌ی \mathbf{V} اختصاص می‌دهد از

$$p(\mathbf{V}) = \sum_{\mathbf{h}} \frac{1}{Z} e^{-E(\mathbf{V}, \mathbf{h})} \quad ; \quad Z = \sum_{\mathbf{V}, \mathbf{h}} e^{-E(\mathbf{V}, \mathbf{h})} \quad (۱۸-۴)$$

بدست می‌آید.

در این ساختار توزیع‌های شرطی برای لایه‌های قابل مشاهده و پنهان به شکل

$$p(v_{ik} = 1 | \mathbf{h}) = \frac{\exp(a_{ik} + \sum_{j=1}^H h_j W_{ijk})}{\sum_{k=1}^K \exp(a_{ik} + \sum_{j=1}^H h_j W_{ijk})} \quad (۱۹-۴)$$

و

$$p(h_j = 1 | \mathbf{V}) = \sigma \left(b_j + \sum_{i=1}^D \sum_{k=1}^K W_{ijk} v_{ik} \right) \quad (۲۰-۴)$$

هستند. همان‌طور که مشاهده می‌گردد توزیع شرطی مربوط به لایه‌ی قابل مشاهده کاملاً با رابطه‌ی ۹-۴ که برای مدل RBM استاندارد تعریف گردید تفاوت دارد، و توزیع مورد استفاده در رابطه‌ی ۱۹-۴ مانند آنچه که در بخش ۷-۲ توضیح داده شد از تابع Softmax استفاده می‌کند. دلیل استفاده از تابع Softmax به جای تابع سیگموید که در مدل RBM استاندارد از آن استفاده می‌شد تغییر در ساختار ورودی و لایه‌ی قابل مشاهده است. با توجه به اینکه هر ویژگی تنها یک مقدار از K مقدار ممکن را می‌تواند داشته باشد، یا به عبارتی در هر ستون ماتریس باینری V در لایه‌ی قابل مشاهده تنها یک ۱ می‌تواند وجود داشته باشد، لذا برای هر ستون در این ماتریس از یک تابع Softmax استفاده می‌گردد و احتمال یک بودن هر یک از این K مقدار حساب می‌گردد و سپس با تولید یک عدد تصادفی با توجه به احتمالات بدست آمده برای هر یک از K مقدار ممکن برای هر ویژگی، مقدار آن ویژگی تعیین می‌گردد. در حقیقت فرمول ۱۹-۴ برای توزیع احتمالی لایه قابل مشاهده این اطمینان را به ما می‌دهد که مقادیر بدست آماده برای هر ویژگی به ازای K مقدار ممکن برای آن، یک توزیع چند احتمالی چندجمله‌ای از درجه‌ی K است.

۴-۴ مدل مولد براساس تعداد کلمات

مدلی که در این بخش به صورت کامل آن را شرح می‌دهیم و گسترش یافته‌ی مدل معرفی شده در بخش ۳-۴ است در واقع همان مدل RS است که در بخش ۵-۳-۳ به صورت مختصر آن را معرفی کردیم. بررسی مدل RS در درک مدل پیشنهادی حائز اهمیت است، چرا که مدل پیشنهادی در این پایان‌نامه بر روی ساختار این مدل سوار شده و گسترش یافته‌ی آن است. مدل RS همان‌طور که در بخش ۵-۳-۳ بیان شد یک مدل مولد احتمالی بر اساس تعداد کلمات است که به صورت بدون نظارت به مدل‌سازی موضوع در داده‌های متنی می‌پردازد.

در بحث مدل‌سازی موضوع پیش از آموزش مدل ابتدا یک بار تمام مجموعه سند پیمایش می‌شود و تمام کلمات متمایز در تمام اسناد مشخص شده و همراه با ایندکس هر کدام ذخیره می‌شوند. به عبارت دیگر قبل از شروع مرحله‌ی آموزش ابتدا یک لغت‌نامه از تمام کلمات متمایز در مجموعه اسناد ساخته می‌شود. پس از پایان این مرحله و ساخت لغت‌نامه به مدل بخش ۳-۴ بر می‌گردیم و مدل معرفی شده در آن جا را در این حالت اندکی بررسی می‌کنیم. در بخش ۳-۴ گفته شد که مقدار هر ویژگی می‌تواند از یک کمینه تا یک بیشینه باشد، حال فرض کنید که بحث مدل‌سازی موضوع بر روی داده‌های متنی است و پس از ساخت لغت‌نامه می‌خواهیم از مدل بخش ۳-۴ استفاده کنیم. در اینجا مقدار هر ویژگی برابر با اندیس یکی از کلمات لغت‌نامه است. هر سند ورودی پس از انجام پیش پردازش‌های لازم به صورت یک دنباله از کلمات تبدیل شده است که هر کدام از این کلمات برابر با یکی از کلمات لغت‌نامه هستند. به این ترتیب در ماتریس ورودی به مدل که یک ماتریس

به اندازه‌ی $K \times D$ می‌باشد K برابر با اندازه لغت‌نامه است و D نشان دهنده‌ی طول سند متنی است. در این حالت برای هر ستون مقدار سطر متناظر با اندیس آن کلمه در لغت‌نامه برابر با ۱ می‌شود و دیگر درایه‌های آن ستون همچنان صفر باقی می‌مانند.

حال فرض کنید که برای مدل کردن داده‌های متنی با استفاده از ساختار معرفی شده در بخش ۳-۴ برای هر سند یک شبکه‌ی RBM جدا می‌سازیم که به تعداد کلمات همان سند دارای واحد Softmax است. در این حالت ورودی دیگر یک ماتریس باینری نخواهد بود و به صورت برداری از تعداد کلمات موجود در آن سند است که می‌توانیم ترتیب را در آن‌ها نادیده بگیریم. رابطه‌ی محاسبه‌ی انرژی در این حالت به شکل

$$E(\mathbf{v}, \mathbf{h}) = - \sum_{j=1}^H \sum_{k=1}^K W_{jk} h_j \hat{v}_k - \sum_{k=1}^K v_k a_k - D \sum_{j=1}^H h_j b_j \quad (21-4)$$

است که در آن $\hat{v}^k = \sum_{i=1}^D v_{ik}$ است. به عبارت دیگر $\hat{\mathbf{v}}$ برداری است با طول K که همان اندازه لغت‌نامه است و از محاسبه‌ی حاصل جمع سطرهای ماتریس باینری ورودی بدست می‌آید. با تغییرات انجام گرفته در لایه‌ی قابل مشاهده و تعویض ساختار ورودی، در ماتریس وزن بین این لایه و لایه‌ی پنهان و همچنین بایاس لایه‌ی قابل مشاهده تغییراتی رخ می‌دهد. در رابطه ۲۱-۴، W ماتریسی با اندازه‌ی $K \times H$ و \mathbf{a} یک بردار با اندازه‌ی $1 \times K$ هستند.

در این مدل که آن را Softmax تکرار شده نامیدیم روابط شرطی محاسبه‌ی لایه‌ی قابل مشاهده و لایه‌ی پنهان به شکل

$$p(v_i = w | \mathbf{h}) = \frac{\exp(a_w + \sum_{j=1}^H h_j W_{wj})}{\sum_{k=1}^K \exp(a_k + \sum_{j=1}^H h_j W_{wj})} \quad (22-4)$$

و

$$p(h_j = 1 | \mathbf{V}) = \sigma \left(Db_j + \sum_{k=1}^K W_{kj} \hat{v}_k \right) \quad (23-4)$$

هستند.

همان‌طور که مشاهده می‌شود در روابط ۲۱-۴ و ۲۳-۴ ترم بایاس برای لایه‌ی پنهان با اندازه سند جاری نیز متناسب است. وجود این تناسب در پیاده‌سازی‌های تجربی و هنگامی که با اسناد با طول‌های متفاوت سر و کار داریم بسیار حیاتی است. در این مدل نیز برای آموزش از الگوریتم CD استفاده می‌شود، رابطه‌ی به روزرسانی نیز همان است که در بخش ۲-۴ و رابطه ۱۶-۴ نشان داده شد با این تفاوت که به جای بردار باینری برای لایه‌ی قابل مشاهده از مقدار آن بر اساس تعداد کلمات استفاده می‌شود. برای مثال در یک مجموعه از N سند

آموزش که به شکل $\{\mathbf{V}_n\}_{n=1}^N$ است رابطه‌ی به روزرسانی برای یک عنصر ماتریس وزن به شکل

$$\frac{1}{N} \sum_{n=1}^N \frac{\partial \log P(\mathbf{V}_n)}{\partial W_{jk}} = E_{P_{data}}[\hat{v}_k h_j] - E_{P_{model}}[\hat{v}_k h_j] \quad (24-4)$$

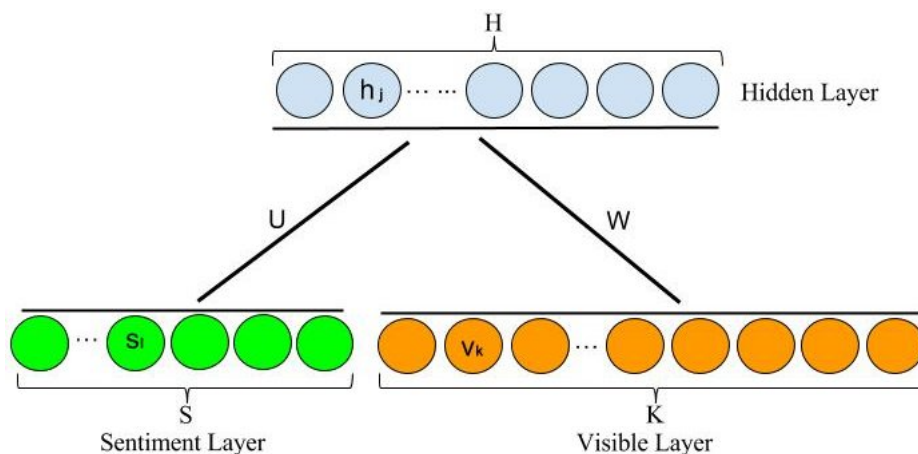
$$\Delta W_{jk} = \alpha (E_{P_{data}}[\hat{v}_k h_j] - E_{P_{model}}[\hat{v}_k h_j]) \quad (25-4)$$

است.

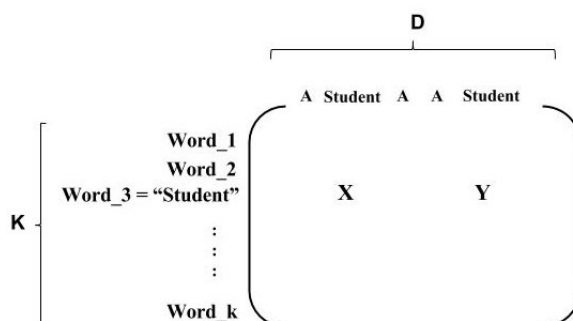
۵-۴ مدل پیشنهادی مولد احتمالی احساس/موضوع

مدل معرفی شده در این پایان‌نامه یک مدل مولد احتمالاتی شبکه عصبی برای مدل‌سازی موضوع و احساس در داده‌های متنی است. این روش گسترش یافته‌ی مدل RS است که در بخش ۴-۴ معرفی شد. در مدل پیشنهادی که در شکل ۳-۴ نشان داده شده است، مشاهده می‌گردد که این روش نیز یک ساختار دو لایه دارد که در سمت لایه‌ی قابل مشاهده‌ی آن یک بردار متناظر با برجسب هر سند که در این پژوهش ما آن را به عنوان بردار متناظر با احساس هر سند تعبیر می‌کنیم به ساختار مدل اضافه شده است. بردار ورودی در این ساختار در قسمت Visible یک بردار با طولی ثابت و به اندازه‌ی اندازه لغت‌نامه یا همان تعداد کلمات متمایز در متن است که در آن تعداد تکرار کلمات مشخص شده است.

در مدل معرفی شده در بخش ۳-۴ که در شکل ۲-۴ نشان داده شده است، هر سند ورودی پس از انجام پیش‌پردازش‌های مورد نیاز و تبدیل به یک دنباله از کلمات به صورت یک ماتریس به مدل وارد می‌شود. در نظر گرفتن یک چنین ساختاری برای هر سند ورودی به این معنی است که جایگاه هر کلمه در متن برای ما مهم است و ترتیب کلمات در هر سند در نظر گرفته می‌شود که این امر موجب بزرگ شدن فضای پارامترهای مساله (سه بعدی شدن ماتریس وزن و دو بعدی شدن ماتریس بایاس برای لایه‌ی قابل مشاهده) و کند شدن فرآیند آموزش می‌گردد. علاوه بر این در بحث مدل‌سازی موضوع حضور و عدم حضور کلمات به همراه فرکانس تکرار آن‌ها دارای اهمیت است نه محل قرار گرفتن هر کلمه در متن، چرا که تمام مدل‌های بررسی شده در این پژوهش و همچنین مدل پیشنهادی بر اساس کیسه کلمات (۲-۱۱) رفتار می‌کنند که در آن ترتیب کلمات در نظر گرفته نمی‌شود. در شکل ۴-۴ یک نمونه از ماتریس ورودی به مدل معرفی شده در بخش ۳-۴ نشان داده شده است. در این مثال سند ورودی به مدل پس از انجام پیش‌پردازش‌های لازم به یک دنباله از پنج کلمه تبدیل شده است، یک کلمه از لغت‌نامه مانند Student در سند ورودی در دو جایگاه مختلف در سند نشان داده شده است، با توجه به ساختار معرفی شده در بخش ۳-۴ و ماتریس‌های وزن و بایاس سه و دو بعدی، نتیجه گرفته می‌شود برای این کلمه دو مقدار متفاوت برای وزن و بایاس آن (یکی برای x و یکی برای y) تعریف می‌شود. لذا متناسب با



شکل ۴-۳: مدل پیشنهادی مولد احتمالی احساس/موضوع با Softmax تکرار شده



شکل ۴-۴: یک سند ورودی با ۵ کلمه در مدل بخش ۴-۳

اینکه کلمه در کجای سند ورودی قرار داشته باشد می‌تواند وزن‌های متفاوتی داشته باشد. این امر موجب بزرگ شدن ماتریس ورودی، طولانی شدن فرایند آموزش و در نظر گرفتن تفاوت برای محل قرار گرفتن هر کلمه در متن می‌گردد. در مدل پیشنهادی در این پژوهش که در ادامه شرح داده می‌شود با تغییر ساختار ورودی از این مشکلات اجتناب می‌گردد.

در مدل پیشنهادی در این پژوهش و همچنین مدل RS هر سند به صورت یک بردار شامل تعداد کلمات به مدل وارد می‌شود. در این ساختار مانند آنچه که در شکل ۴-۳ نشان داده شده است طول بردار ورودی برابر با اندازه لغت‌نامه یا همان تعداد کلمات متمایز در مجموعه سند در نظر گرفته می‌شود که درایه‌های آن تعداد تکرار هر کلمه از لغت‌نامه در سند جاری را نشان می‌دهند. در این حالت در واقع وزن‌ها برای هر کلمه به اشتراک گذاشته می‌شوند فارغ از اینکه این کلمه در کجای سند ورودی قرار دارد. به طور مثال برای کلمه‌ی n ام لغت‌نامه یک وزن و یک بایاس تعریف می‌گردد و این کلمه در هر جای سند ورودی قرار داشته باشد وزنش تغییری نخواهد کرد و در فرایند آموزش تنها یک وزن و یک بایاس برای هر کلمه یاد گرفته می‌شود. با توجه به ساختار مدل که در شکل ۴-۳ نشان داده شده است، برای هر سند متنی برداری با اندازه لغت‌نامه که شامل تعداد تکرار

هر کلمه از لغتنامه در آن سند است و همچنین یک بردار باینری که نشان دهنده احساس سند جاری است به عنوان ورودی به شبکه وارد می‌شوند، و توزیع‌های موجود بر روی کلمات مختلف در هر موضوع و همچنین احساس مرتبط با آن‌ها توسط مدل در لایه‌ی پنهان استخراج می‌شوند.

مدل پیشنهادی در این پژوهش (شکل ۴-۳) یک مدل نظارت شده است که برای هر سند با کمینه کردن یک تابع انرژی که به شکل

$$E(\mathbf{V}, \mathbf{s}, \mathbf{h}) = - \sum_{j=1}^H \sum_{k=1}^K W_{kj} h_j \hat{v}_k - \sum_{j=1}^H \sum_{l=1}^S U_{lj} h_j s_l - \sum_{k=1}^K v_k a_k - \sum_{l=1}^S s_l c_l - D \sum_{j=1}^H h_j b_j \quad (۲۶-۴)$$

تعریف می‌شود، به مدل‌سازی احساس و موضوع به صورت مشترک در داده‌های متنی می‌پردازد. با اضافه شدن یک لایه‌ی جدید در این مدل به همراه ماتریس وزن و همچنین بردار بایاس همراه با آن، روابط موجود نسبت به ساختارهای پیشین تغییر می‌کند. همان‌طور که در رابطه‌ی ۴-۲۶ مشاهده می‌شود، برای محاسبه‌ی انرژی در مدل پیشنهادی ترم‌های مربوط به وزن و بایاس لایه‌ی احساس نیز در بدست آوردن مقدار نهایی مشارکت دارند. پس از محاسبه‌ی مقدار انرژی به کمک رابطه‌ی ۴-۲۶، با استفاده از

$$p(\mathbf{v}, \mathbf{s}, \mathbf{h}) = \frac{1}{Z} e^{-E(\mathbf{v}, \mathbf{s}, \mathbf{h})} \Rightarrow p(\mathbf{v}, \mathbf{s}) = \frac{1}{Z} \sum_{\mathbf{h}} e^{-E(\mathbf{v}, \mathbf{s}, \mathbf{h})}, Z = \sum_{\mathbf{v}} \sum_{\mathbf{s}} \sum_{\mathbf{h}} e^{-E(\mathbf{v}, \mathbf{s}, \mathbf{h})} \quad (۲۷-۴)$$

احتمالی که مدل به هر سند و لایه‌ی احساس همراه با آن اختصاص می‌دهد، محاسبه می‌گردد. برای آموزش این مدل و بروزرسانی پارامترهای شبکه که شامل ماتریس‌های وزن بین لایه‌ی قابل مشاهده و پنهان و همچنین لایه‌ی احساس و پنهان هستند و همچنین بایاس‌های هر سه لایه از الگوریتم CD که در بخش ۲-۱۰ معرفی شده به شکل

$$\Delta \theta = \alpha (E_{P_{data}}[\theta] - E_{P_{model}}[\theta]) \Rightarrow \theta_{t+1} = \theta_t + \Delta \theta \quad (۲۸-۴)$$

استفاده می‌شود. در رابطه ۴-۲۶، $\theta = \{W, U, \mathbf{a}, \mathbf{b}, \mathbf{c}\}$ مجموعه پارامترهای مدل است که در آن $W_{K \times H}$ ماتریس وزن بین بردار $Visible$ و لایه‌ی $Hidden$ ، $U_{S \times H}$ ماتریس وزن بین لایه‌ی $Sentiment$ و لایه‌ی $Hidden$ و \mathbf{a} ، \mathbf{b} و \mathbf{c} به ترتیب بردارهای بایاس لایه‌ی $Visible$ ، $Hidden$ و $Sentiment$ می‌باشند. لازم به ذکر است که K و H مانند آنچه در بخش‌های پیشین ذکر کردیم به ترتیب اندازه لغتنامه و طول لایه‌ی پنهان هستند و S به عنوان تعداد احساس موجود یا اندازه بردار $Sentiment$ تعریف می‌شود.

در مدل پیشنهادی روابط شرطی برای محاسبه‌ی هر یک از لایه‌های $Visible$ ، $Sentiment$ و $Hidden$ به

شکل

$$p(v_i = w | \mathbf{h}) = \frac{\exp(a_w + \sum_{j=1}^H W_{wj} h_j)}{\sum_{k=1}^K \exp(a_k + \sum_{j=1}^H W_{kj} h_j)} \quad (29-4)$$

$$p(s_l = 1 | \mathbf{h}) = \frac{\exp(c_l + \sum_{j=1}^H U_{lj} h_j)}{\sum_{l=1}^S \exp(c_l + \sum_{j=1}^H U_{lj} h_j)} \quad (30-4)$$

$$p(h_j = 1 | \mathbf{v}, \mathbf{s}) = \sigma \left(Db_j + \sum_{k=1}^K W_{kj} \hat{v}_k + \sum_{l=1}^S U_{lj} s_l \right) \quad (31-4)$$

هستند. در اینجا چون مقدار لایه‌ی *Hidden* به هر دو مقدار لایه‌ی *Visible* و *Sentiment* وابسته است، لذا مشاهده می‌شود که در رابطه‌ی ۳۱-۴ برای مقدار لایه‌ی *Hidden* از یک توزیع شرطی که وابسته به هر دو مقدار لایه‌های *Visible* و *Sentiment* است نمونه گرفته می‌شود. اما با توجه به اینکه با داشتن مقدار لایه‌ی *Hidden* بردارهای *Visible* و *Sentiment* از یکدیگر مستقل شرطی هستند لذا در روابط ۲۹-۴ و ۳۰-۴ مقدار این دو بردار از یک توزیع شرطی که تنها به مقدار بردار *Hidden* وابسته است نمونه گرفته می‌شوند.

با توجه به خصوصیات بیان شده برای تابع *Softmax* در بخش ۷-۲ و توجه به این امر که خروجی برای این تابع متناظر با مقادیر یک توزیع احتمالی چندجمله‌ای است، لذا همان‌طور که مشاهده می‌شود، در روابط ۲۹-۴ و ۳۰-۴ برای محاسبه‌ی مقادیر لایه‌های *Visible* و *Sentiment* از یک تابع *Softmax* استفاده می‌گردد. در فرآیند آموزش با استفاده از الگوریتم CD برای بدست آوردن مقدار بازسازی^۱ شده از لایه‌ی *Visible* مشروط به بردار *Hidden* از رابطه‌ی ۲۹-۴ که به صورت *Softmax* است، استفاده می‌شود. در واقع دلیل اینکه این رابطه و رابطه‌ی ۳۰-۴ برای لایه‌ی *Sentiment* به فرم تابع *Softmax* هستند همین امر است، که پس از محاسبه‌ی مقادیر این لایه‌ها مشروط به بردار *Hidden* نیاز به تولید نمونه و نمونه‌برداری از این مقادیر بدست آمده داریم. در نتیجه استفاده از تابع *Softmax* برای ما تضمین می‌کند که مقادیر محاسبه شده برای این دو بردار یک توزیع احتمالی چندجمله‌ای خواهد بود که می‌توان به راحتی از آن نمونه تولید کرد.

۴-۶ نتیجه‌گیری

در این بخش کلیت نظری مدل پیشنهادی توضیح داده شد. در ابتدا مدل RBM به عنوان ساختار پایه برای مدل پیشنهادی در این پایان‌نامه مطرح و به صورت کامل بررسی گردید. پس از آن دو روش گسترش یافته از مدل RBM که مدل پیشنهادی در این پایان‌نامه تعمیم یافته‌ی آن‌ها برای مدل‌سازی احساس و موضوع است به صورت

^۱Reconstruct

دقیق مورد بررسی قرار گرفتند. سپس مدل پیشنهادی در این پایان‌نامه به همراه روابط ریاضی و مؤلفه‌های موجود در آن به صورت کامل توضیح و اثبات گردیدند.

در فصل بعدی جزییات بیشتری در مورد شبیه‌سازی مدل پیشنهادی و نتایج حاصل شده از آن تشریح می‌شود.

فصل پنجم

شبیه‌سازی و ارزیابی مدل پیشنهادی

۵-۱ مقدمه

در فصل پیشین مدل پیشنهادی در این پژوهش به طور کامل معرفی گردید و با بخش‌ها و روابط موجود در آن کاملاً آشنا شدیم. در این فصل رویکرد معرفی شده در این پایان‌نامه مورد ارزیابی و آزمایش قرار می‌گیرد و نتایج بدست آمده در آزمایش‌های گوناگون را مورد تحلیل و بررسی قرار می‌دهیم. در این فصل ابتدا پیش‌پردازش‌های لازم و مطرح در بحث NLP را که از آن‌ها برای آماده‌سازی پایگاه داده استفاده می‌کنیم شرح داده و سپس پایگاه داده‌ی مورد استفاده در این فصل را معرفی کرده و خصوصیات آن را در سه حالت مختلف بیان می‌کنیم، سپس یک معیار معروف به نام سرگشتگی^۲ که از آن برای ارزیابی مدل‌های احتمالی مولد استفاده می‌شود و در ادامه از آن برای ارزیابی مدل پیشنهادی استفاده می‌کنیم را تعریف می‌کنیم. در بخش‌های بعدی آزمایشات انجام شده بر روی پایگاه داده‌ی معرفی شده را به تفصیل شرح می‌دهیم و نتایج بدست آمده را تحلیل می‌کنیم.

^۲Perplexity

۵-۲ پیش‌پردازش‌های متنی در پردازش زبان طبیعی

در مباحث مربوط به NLP مخصوصاً روش‌ها و رویکردهایی که با داده‌های متنی سر و کار دارند، پیش از اینکه داده‌ها به مدل وارد شوند بر روی آن‌ها پیش‌پردازش‌های تاثیر گذاری انجام می‌شود که ما نیز برای آماده‌سازی داده‌ها جهت ورود به مدل این کار را انجام داده و پس از تبدیل داده‌های متنی به شکلی استاندارد آن‌ها را به مدل وارد کردیم. چهار پیش‌پردازش لازم جهت آماده سازی داده‌های متنی به ترتیب اجرا بر روی پایگاه داده به صورت:

۱. علامت‌گذاری و حذف کارکترهای بی‌معنی^۱

۲. ریشه یابی لغوی^۲

۳. ریشه یابی نحوی^۳

۴. حذف کلمات توقف^۴

می باشند که در ادامه به تعریف آنها می‌پردازیم.

۵-۲-۱ علامت‌گذاری و حذف کاراکترهای بی‌معنی

در این مرحله هر جمله به کلمات تشکیل دهنده‌ی آن شکسته می‌شود و سپس تمام کاراکترهای اضافه‌ای که در واژگان زبان وجود نداشته باشند از جمله حذف می‌گردند. برای مثال در این مرحله نقطه گذاری‌ها، پرانتزها، کاراکترهای خاص مانند @, * و غیره از تمام اسناد در پایگاه داده حذف می‌شوند.

۵-۲-۲ ریشه یابی لغوی

ریشه یابی لغوی به معنی تبدیل حالت‌های مختلف یک کلمه که در متن وجود دارند به یک صورت واحد و یکتا است. به طور مثال در زبان انگلیسی می‌توان به مواردی نظیر حذف ing از پایان کلمات یا حذف 's مالکیت از انتهای اسامی اشاره کرد. برای درک بهتر به مثال زیر توجه کنید که در آن تمام حالت‌های سمت چپ در فرآیند پیش‌پردازش ریشه یابی لغوی به حالت سمت راست تبدیل می‌شوند.

$car, cars, car's, cars' \rightarrow car$

¹Tokenization and Remove Meaningless Characters

²Stemming

³Lemmatization

⁴Remove Stop Words

۵-۲-۳ ریشه یابی نحوی

علی‌رغم نتایج یکسان برای ریشه یابی لغوی و نحوی در بسیاری از حالت‌ها، اما تفاوت فراوانی بین این دو پیش‌پردازش وجود دارد. در بحث ریشه یابی نحوی ما به دنبال یافتن ریشه‌ی افعال و کلمات موجود در متن هستیم و بسیاری از کلمات و افعال به شکل مصدری خود بازگردانده می‌شوند. ریشه یابی لغوی برای هر کلمه به صورت جداگانه و بدون در نظر گرفتن مفهوم متن انجام می‌شود، اما در حالت نحوی با توجه به مفهوم، کلماتی که شکل یکسانی دارند اما معنی آن‌ها متفاوت است به مصدر های مختلفی تبدیل می‌شوند. به طور مثال در زبان انگلیسی برای کلمه‌ای مانند Walking نتیجه‌ی حاصل از هر دو ریشه یابی نحوی و لغوی کلمه‌ی Walk است. اما برای سه فعل کمکی مانند *am, is, are* با انجام ریشه یابی لغوی تغییری در آن‌ها اتفاق نمی‌افتد، اما با ریشه یابی نحوی هر سه کلمه به حالت *be* تغییر شکل می‌دهند.

“Walking” with Stemming → Walk “Walking” with Lemmatization → Walk
 “Am, Is, Are” with Stemming → Am, Is, Are “Am, Is, Are” with Lemmatization → be

۵-۲-۴ حذف کلمات توقف

در هر زبانی لغات بسیاری هستند که آن‌ها را کلمات عمومی یا کلمات توقف آن زبان تعریف می‌کنند. این کلمات به صورت گسترده و فراوان در متن یافت می‌شوند و هیچ‌گونه بار اطلاعاتی با خود به همراه ندارند. حذف این کلمات از داده‌های متنی علاوه بر کوچک کردن اندازه داده‌ی ورودی به مدل منجر به بهبود کیفیت نتایج شده و کارایی مدل را افزایش می‌دهد. در ادامه چند نمونه از این کلمات برای زبان انگلیسی آورده شده است.

me, my, myself, ourselves, can, will, just, ...

۵-۳ پایگاه داده‌ی بازبینی فیلم

پایگاه داده‌ی بازبینی فیلم^۱ (MR) پس از استفاده در کار Pang و همکاران [۲۶] تبدیل به یک معیار در بحث مدل‌سازی احساس و همچنین ارزیابی مدل‌های پیشنهاد شده گردیده است [۶]. نسخه ۲۲ از این پایگاه داده که ما در آزمایش‌های خود از آن استفاده می‌کنیم شامل ۱۰۰۰ بازبینی مثبت از فیلم‌های مختلف و ۱۰۰۰ بازبینی منفی است. این بازبینی‌ها از سایت پایگاه داده‌ی اینترنتی فیلم^۲ (IMDB) جمع‌آوری شده‌اند. میانگین طول هر بازبینی در این پایگاه داده ۳۰ جمله است که بر روی آن پیش‌پردازش معرفی شده در بخش‌های ۵-۲-۱ تا

¹Movie Review

²Available at <http://www.cs.cornell.edu/people/pabo/movie-review-data>

³Internet Movie Database

۲-۲-۵ را انجام می‌دهیم و در ادامه مراحل آن را شرح می‌دهیم.

۴-۵ پایگاه داده‌ی ۲۰ گروه خبری

پایگاه داده‌ی ۲۰ گروه خبری^۱ (۲۰NG) یکی از دیتاست‌های معروف در بحث مدل‌سازی موضوع است. این پایگاه داده شامل ۱۸۷۸۶ سند متنی است که از مخازن گروه‌های خبری Usenet جمع‌آوری شده‌اند. این مجموعه سند به ۲۰ گروه خبری مختلف تقسیم می‌شود که هر کدام از این ۲۰ گروه مربوط به یک موضوع خاص هستند. از مجموع ۱۸۷۸۶ سند موجود در این پایگاه داده، ۱۱۲۸۴ سند برای مجموعه‌ی آموزش و ۷۵۰۲ سند برای مجموعه‌ی تست در نظر گرفته می‌شوند. پس از انجام پیش‌پردازش‌های گفته شده در بخش ۲-۵، ۲۰۰۰ کلمه‌ای که بیشترین تکرار را دارند جدا شده و به عنوان لغت‌نامه برای این پایگاه داده در نظر گرفته می‌شوند. در بخش‌های بعدی از لغت‌نامه‌ی این پایگاه داده برای ارزیابی‌های بسیاری استفاده می‌کنیم. همچنین در بخش ۱۲-۵ از این پایگاه داده استفاده و نتایج بدست آمده از بازیابی اطلاعات بر روی آن را گزارش می‌کنیم.

۵-۵ پایگاه داده‌ی احساس چند دامنه

پایگاه داده‌ی احساس چند دامنه^۲ (MDS) اولین بار توسط Blitzer و همکاران [۲۷] در سال ۲۰۰۷ مورد استفاده قرار گرفت. این پایگاه داده شامل بازیابی‌های نوشته شده در مورد چهار نوع مختلف از محصولات سایت آمازون است که جمع‌آوری شده‌اند. بازیابی‌های موجود در این دیتاست مربوط به چهار گروه کتاب، دی‌وی‌دی، وسایل الکترونیکی و وسایل آشپزخانه هستند. برای هر یک از این چهار دسته ۱۰۰۰ بازیابی مثبت و ۱۰۰۰ بازیابی منفی در MDS وجود دارد. در بخش ۶-۵ از ترکیب MDS با MR که در بخش ۳-۵ معرفی شد، برای ساخت یک پایگاه داده‌ی بزرگتر استفاده می‌شود که برای ارزیابی مدل پیشنهادی در فرآیند بازیابی اطلاعات در بخش ۱۲-۵ از آن استفاده می‌کنیم.

۶-۵ آماده‌سازی پایگاه داده

در این قسمت مراحل آماده‌سازی پایگاه داده MR برای استفاده در بخش‌های آینده به صورت کامل توضیح داده می‌شود. پس از انجام تمام پیش‌پردازش‌های گفته شده در بخش ۲-۵ به همان ترتیب گفته شده بر روی داده‌های آموزش و آزمون در پایگاه داده‌ی MR، هر سند متنی به دنباله‌ای از کلمات تبدیل می‌شود. مرحله‌ی

¹News Groups, Available at <http://people.csail.mit.edu/jrennie/20Newsgroups>

²Multi Domain Sentiment, Available at <http://www.cs.jhu.edu/mdredze/datasets/sentiment/index2.html>

Data Set	Dictionary Size	Num of Train	Num of Test	Avg Docs Length	Std Deviation
Movie Review	2000	1000	1000	90.18	40.23
Movie Review	10000	1000	1000	186.35	81.33
Movie Review	24916	1000	1000	299.75	126.51

جدول ۵-۱: اطلاعات آماری پایگاه داده‌ی Movie Review

بعدی ساخت لغت‌نامه و تبدیل پایگاه داده به فایل lib-svm برای ورود به مدل است. در این قسمت علاوه بر استفاده از دو لغت‌نامه معروف در بحث مدل‌سازی موضوع، یک لغت‌نامه نیز از داده‌های پیش‌پردازش شده ساخته شد. برای ساخت این لغت‌نامه‌ی واژگان تمام اسناد را به صورت کامل پیمایش کردیم تا کلمات متمایز در آن‌ها مشخص گردند. مشاهده گردید که تعداد کلمات متمایز در این حالت ۲۴۹۱۶ عدد است. در نتیجه اندازه لغت‌نامه در این حالت برابر با ۲۴۹۱۶ در نظر گرفته شد. همان‌طور که بیان گردید از دو لغت‌نامه دیگر با اندازه‌های ۲۰۰۰ و ۱۰۰۰۰ کلمه‌ی متمایز برای ساخت فایل lib-svm برای پایگاه داده نیز استفاده کردیم. این دو لغت‌نامه به ترتیب مربوط به دو پایگاه داده‌ی NG۲۰ و توده‌ی اسناد رویت‌ر نسخه ۱۱ (RCV۱) هستند که از پایگاه داده‌های معیار در بحث مدل‌سازی موضوعی هستند. اطلاعات آماری بدست آمده پس از انجام مراحل گفته شده در جدول ۵-۱ نشان داده شده است.

همان‌طور که بیان شد پایگاه داده‌ی MR شامل ۲۰۰۰ سند است که ۱۰۰۰ عدد از این اسناد مثبت و ۱۰۰۰ سند باقی مانده دارای برچسب احساس منفی هستند. مانند آنچه که در مدل JST [۶] استفاده شده است در اینجا ما نیز این پایگاه داده را به دو دسته‌ی آموزش و آزمون تقسیم می‌کنیم. تعداد اسناد در هر یک از این دو گروه ۱۰۰۰ است که ۵۰۰ عدد از آن‌ها مثبت و ۵۰۰ تای دیگر دارای برچسب منفی هستند. لازم به ذکر است فایل ورودی به مدل یک فایل به صورت lib-svm است. هر سند متنی در این فایل متناظر با یک سطر است که فرمت آن به شکل:

label <ID:Count> <ID:Count> <ID:Count> ...

است. ID در اینجا نشان دهنده‌ی ایندکس هر یک از کلمات در لغت‌نامه و Count نشان دهنده تعداد تکرار آن کلمه در متن جاری است. همچنین در ابتدای هر سطر یک عدد تنها که مقدار آن ۱ یا ۲ است وجود دارد که نشان دهنده‌ی برچسب احساس آن سند است. مقدار ۱ نمایانگر احساس مثبت و مقدار ۲ نمایانگر احساس منفی در این حالت است. در شکل ۵-۱ نمونه‌ای از فایل lib-svm برای یک سند نشان داده شده است. همان‌طور که ملاحظه می‌شود با توجه به اینکه عدد اول برای این سند ۲ است، لذا این فایل متعلق به یک سند با برچسب احساس منفی است.

¹ Reuters Corpus Volume 1, Available at <http://trec.nist.gov/data/reuters/reuters.html>

```
2 7:1 11:3 14:1 17:1 24:1 34:2 60:1 69:1 72:1 74:3 96:1 99:1 102:1 123:1 130:1
131:1 193:2 197:1 204:1 205:1 210:1 280:2 293:1 294:1 296:1 308:1 342:1 353:1
355:4 362:1 372:1 378:1 443:2 470:1 492:1 496:2 501:4 576:1 595:1 596:3 599:1
626:3 719:1 747:1 794:1 802:1 838:1 922:1 937:1 938:1 1007:2 1071:1 1226:2 1235:1
1265:1 1294:1 1307:1 1320:1 1341:1 1551:1 1716:1 1748:1 1791:1 1969:1
```

شکل ۵-۱: نمونه‌ای از یک سند منفی در فایل lib-svm

۵-۲ لغت‌نامه‌ی احساس

منظور از لغت‌نامه‌ی احساس یک لغت‌نامه عمومی از پیش ساخته شده است که در آن به ازای هر کلمه برای هر یک از برچسب‌های احساس مثبت، منفی و بی‌طرف وزنی بین ۰ تا ۱ وجود دارد به گونه‌ای که مجموع این مقادیر برای هر کلمه برابر با ۱ است. به عبارت دیگر این لغت‌نامه شامل تعدادی کلمه است که به هیچ دامنه‌ی خاصی وابسته نیستند و برچسب احساسی برای آنها مشخص است. منظور از اینکه این کلمات به هیچ دامنه‌ی خاصی وابسته نیستند و مستقل از دامنه هستند این امر است که به طور مثال در یک موضوع سینمایی کلمه‌ای مثل “پیچیده” برای یک فیلم‌نامه می‌تواند احساسی مثبت به همراه داشته باشد، اما همین کلمه در موضوع علمی بار مثبتی به همراه خود ندارد. از این رو بعضی از کلمات از نظر احساسی وابسته به دامنه‌ای هستند که در آن بحث می‌شود و در زمینه‌های مختلف می‌توانند بار احساسی متفاوتی داشته باشند. اما لغت‌نامه‌ی احساس از کلماتی ساخته می‌شود که در تمام دامنه‌ها و زمینه‌ها بار احساسی که به همراه خود دارند ثابت است. به طور مثال کلمه‌ای مانند “good” همیشه یک مفهوم مثبت و کلمه‌ای مانند “bad” همه‌جا یک مفهوم منفی را می‌رساند.

در بخش‌های بعدی در یکی از آزمایش‌های طراحی شده برای مدل پیشنهادی در این پژوهش از یک لغت‌نامه‌ی احساس به نام MPQA^۱ استفاده می‌کنیم. این لغت‌نامه احساسی شامل ۴۰۵۳ کلمه است که در مقابل هر لغت یک بردار ۳ تایی وجود دارد که عدد اول نشان دهنده‌ی وزن بی‌طرفی، عدد دوم نشان دهنده‌ی وزن مثبت و عدد سوم نشان دهنده‌ی وزن منفی برای آن لغت است. در مجموع در این لغت‌نامه ۱۵۱۱ کلمه‌ی مثبت و ۲۵۴۲ کلمه‌ی منفی وجود دارد. شکل ۵-۲ نمونه‌ای از کلمات این لغت‌نامه را نشان می‌دهد.

great	0.05	0.9	0.05
bad	0.05	0.05	0.9
destroy	0.05	0.05	0.9
cry	0.05	0.05	0.9
beauty	0.05	0.9	0.05
wonder	0.05	0.9	0.05

شکل ۵-۲: نمونه‌ای از لغت‌نامه احساس MPQA شامل ۶ کلمه

^۱ Available at <http://mpqa.cs.pitt.edu/>

۵-۸ جزئیات آموزش مدل پیشنهادی

رویکرد پیشنهادی در این پژوهش یک مدل احتمالی مولد نظارت شده برای مدل‌سازی همزمان احساس و موضوع در داده‌های متنی است که در بخش ۴-۵ به صورت کامل آن را معرفی کردیم. در پایان‌نامه برای پیاده‌سازی از زبان برنامه‌نویسی پایتون نسخه ۲,۷ در محیط سیستم عامل لینوکس استفاده شده است. برای پیاده‌سازی ابتدا مدل RS براساس آنچه که در [۲] بیان شده است شبیه‌سازی گردید، و پس از ارزیابی و حصول اطمینان از صحت مدل پیاده‌سازی شده با بررسی نتایج حاصل از شبیه‌سازی انجام شده با نتایج موجود در [۳] شبیه‌سازی انجام شده را به مدل پیشنهادی در این پژوهش گسترش دادیم.

براساس آنچه که در بخش ۵-۶ بیان کردیم، از پایگاه داده‌ی MR در سه حالت مختلف به عنوان داده‌ی ورودی به مدل برای فرآیند آموزش و آزمون استفاده کردیم و نتایج بدست آمده از آزمایشات انجام شده را در بخش‌های بعدی شرح می‌دهیم. برای آموزش مدل در سه حالات موجود، یعنی با استفاده از دیکشنری‌های با سائز ۲۰۰۰، ۱۰۰۰۰ و ۲۴۹۱۶ ما از الگوریتم CD با مرتبه‌ی ۱ استفاده کردیم. به این معنی که در مرحله بازسازی داده تنها یک مرحله از این الگوریتم اجرا شده و جهت گرادیان را همان‌طور که آقای Hinton در [۱۰] اثبات کرده‌اند تنها با همین یک مرحله تخمین می‌زنیم. همچنین در هر سه حالت، مدل را به ازای ۱۰۰۰ تکرار^۱ بر روی کل داده‌های آموزش با Batch سائز ۱ آموزش دادیم و نتایج بدست آمده برای هر حالت را در ۲ مرحله، یکی در تکرار ۲۰۰ام و یکی در زمان پایان فرآیند آموزش (تکرار ۱۰۰۰ام) ثبت کردیم. پارامتر دیگری که در آموزش مدل دخیل است تعداد واحدهای لایه‌ی Hidden یا همان تعداد موضوع‌ها است که می‌توانند متغیر باشند. در استفاده از هر ۳ دیکشنری، رویکرد پیشنهادی و همچنین مدل RS را به ازای $h = \{5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90\}$ آموزش دادیم و نتایج بدست آمده برای حالت‌های مختلف را در ادامه در آزمایش‌های مختلف گزارش می‌کنیم. برای تمام حالت‌ها از مقدار $alfa = 0.001$ برای ضریب یادگیری استفاده کردیم. پارامترهای $W, U, \mathbf{a}, \mathbf{c}$ که به ترتیب وزن بین لایه‌ی Visible و Hidden، وزن بین لایه‌ی Sentiment و Hidden، بایاس لایه‌ی Visible و بایاس لایه‌ی Sentiment هستند را با استفاده از مقادیری که به صورت تصادفی از یک توزیع گوسی با میانگین ۰ و واریانس ۱ بدست آوردیم، مقدار دهی کردیم. همچنین مقدار اولیه‌ی برای بایاس لایه‌ی Hidden که آن را با \mathbf{b} نشان دادیم را برابر صفر قرار دادیم.

^۱ Epoch, Iteration

۵-۹ مدل سازی اسناد و ارزیابی به عنوان یک مدل مولد

در این بخش رویکرد پیشنهادی خودمان را به عنوان یک مدل مولد احتمالاتی با مدل RS در تخمین احتمال برای مشاهدهی سندهای پایگاه دادههای آموزش و تست با استفاده از هر سه لغت نامه مورد ارزیابی قرار داده و با تحلیل نتایج بدست آمده نشان می دهیم که روش پیشنهادی نسبت به روش RS یک روش بهتر در تخمین احتمال برای سندهای دیده نشده و آزمون است.

برای ارزیابی احتمال محاسبه شده برای مشاهدهی اسناد در فرآیند مدل سازی مجموعه سند، همان طور که در بخش ۵-۱ گفته شد از یک معیار به نام سرگشتگی استفاده می شود [۱]. در مباحث مربوط به NLP معیار سرگشتگی پارامتری است که از آن برای مقایسهی مدل های احتمالاتی مختلف استفاده می شود [۱]. در فرآیند مدل سازی اسناد ما به دنبال تخصیص بالاترین درست نمایی به هر سند هستیم و با استفاده از معیار سرگشتگی این مقدار درست نمایی محاسبه شده برای هر سند مورد ارزیابی قرار می گیرد. با توجه به فرمول محاسبهی مقدار سرگشتگی،

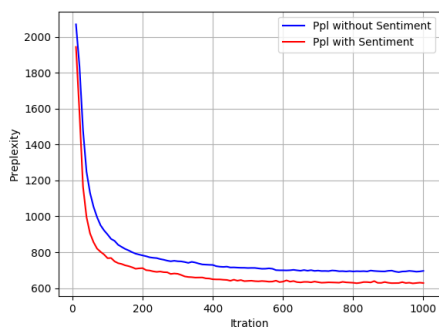
$$Perplexity = \exp \left(- \frac{\sum_{n=1}^N \log p(\mathbf{v}_n)}{\sum_{n=1}^N D_n} \right) \quad (۱-۵)$$

می توان گفت که مقدار این معیار برابر است با معکوس میانگین درست نمایی بدست آماده برای هر سند در مقیاس لگاریتمی به ازای تمام کلمات مجموعه اسناد [۱]. در یک فرآیند مدل سازی و با استفاده از یک مدل احتمالی مناسب مقدار سرگشتگی باید به صورت پیوسته و یکنوا کاهش یابد و مدلی که مقدار سرگشتگی کمتری بر روی پایگاه داده آزمون داشته باشد در بحث مدل سازی اسناد به عنوان مدل بهتری شناخته می شود [۱].

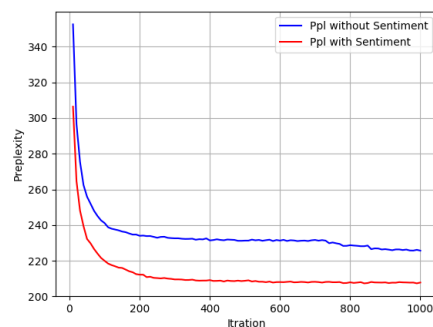
در شکل ۵-۳ قسمت های ۵-۳ تا ۵-۳ ج نمودار تغییرات سرگشتگی در فرآیند آموزش برای مدل پیشنهادی و مدل RS در حالت های مختلف از پایگاه دادهی MR نشان داده شده است. همان طور که مشاهده می گردد در هر ۳ نمودار رویکرد پیشنهادی در این پژوهش که یک مدل مشترک احساس موضوع است نسبت به مدل RS که یک رویکرد موضوعی است با کاهش بهتری در مقدار سرگشتگی همراه است.

برای رسم نمودارهای شکل ۵-۳ برای هر دو مدل پیشنهادی و مدل RS در هر سه حالت مختلف پایگاه داده، در پایان هر مرحلهی آموزش مقدار سرگشتگی با استفاده از رابطهی ۵-۱ برای تمام پایگاه دادهی آموزش محاسبه شده و به ازای هر ۱۰ مرحله از آن میانگین گرفته شده است و در انتها با استفاده از این مقادیر نمودارهای حاصل برای مراحل مختلف آموزش رسم گردیده اند.

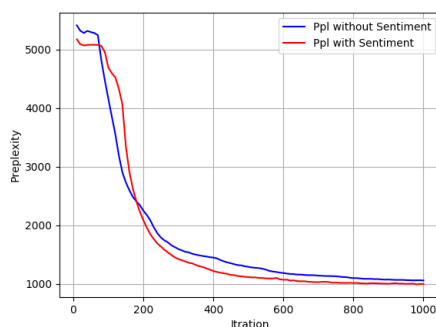
برای هر سه حالت می توان مشاهده کرد که مقدار افت سرگشتگی در ابتدای فرایند آموزش نسبت به مراحل پایانی با سرعت بیشتری همراه بوده است، به صورتی که از تکرار ۲۰۰ام تا به انتها مقدار سرگشتگی با تغییرات



(ب) با استفاده از لغت‌نامه با اندازه ۱۰۰۰۰



(آ) با استفاده از لغت‌نامه با اندازه ۲۰۰۰



(ج) با استفاده از لغت‌نامه با اندازه ۲۴۹۱۶

شکل ۵-۳: ارزیابی تغییرات سرگشتگی در فرآیند آموزش بر روی پایگاه داده‌ی MR برای مدل پیشنهادی و مدل RS

آنچنانی همراه نبوده است. مشاهده‌ی این ویژگی در فرایند آموزش موجب گردید که ما هر دو مدل را برای هر سه حالت مختلف پایگاه داده به ازای دو مقدار ۲۰۰ و ۱۰۰۰ چرخه، آموزش داده و از نتایج بدست آمده برای تست مدل بر روی پایگاه داده آزمون استفاده کنیم.

با دقت در نمودارهای شکل ۵-۳ می‌توان نتیجه گرفت که با اضافه کردن و در نظر گرفتن احساس و ساخت یک مدل مشترک احتمالاتی مولد، مانند آنچه که در این پژوهش انجام دادیم، در مرحله‌ی آموزش برای مدل‌سازی اسناد مقدار سرگشتگی با افت بیشتری همراه می‌شود و در نتیجه روش احتمالاتی مناسب‌تری برای مدل‌سازی اسناد ساخته می‌شود.

مقادیر محاسبه شده برای سرگشتگی که در جدول ۵-۲ نشان داده شده است نیز دلیلی بر اثبات ادعای ما نسبت به بهتر بودن رویکرد پیشنهادی در فرآیند مدل‌سازی به عنوان یک مدل مولد است. در جدول ۵-۲ مقدار سرگشتگی برای داده‌های تست در پایگاه داده‌ی MR به ازای هر ۳ دیکشنری مورد استفاده و ۲ تکرار ۲۰۰ و ۱۰۰۰ برای هر کدام محاسبه شده است.

مقادیر بدست آمده برای سرگشتگی در حالت‌های مختلف در جدول ۵-۲ نشان می‌دهد که در حالت‌های استفاده از ۲ لغت‌نامه ۲۰۰۰ و ۱۰۰۰ تاایی تفاوت مقدار محاسبه شده برای سرگشتگی در مدل‌هایی که

TestSet Type	Num of Docs	Num of Epoch	Ppl without Sentiment	Ppl with Sentiment
MR by 2000	1000	200	400.77	393.69
MR by 2000	1000	1000	423.89	406.74
MR by 10000	1000	200	1553.52	1529.42
MR by 10000	1000	1000	2028.69	1871.57
MR by 24916	1000	200	4237.65	3898.67
MR by 24916	1000	1000	5842.39	5824.97

جدول ۵-۲: تخمین سرگشتگی برای پایگاه داده‌ی Movie Review با استفاده از مدل پیشنهادی

۱۰۰۰ مرحله آموزش دیده‌اند بیشتر است از مدل‌هایی که ۲۰۰ مرحله آموزش دیده‌اند. اما در استفاده از لغت‌نامه‌ی ۲۴۹۱۶ تایی شرایط برعکس است و تفاوت در حالتی که هر دو مدل ۲۰۰ مرحله آموزش دیده‌اند بیشتر از زمانی است که مدل‌ها ۱۰۰۰ مرحله آموزش دیده‌اند. در اینجا به نمودارهای شکل ۳-۵ برمی‌گردیم و با مقایسه‌ی نمودار ۳-۵ ج با نمودارهای ۳-۵ آ و ۳-۵ ب مشاهده می‌کنیم که نمودار مربوط به حالت ۲۴۹۱۶ تایی دارای نوسانات بیشتری نسبت به ۲ حالت دیگر است. به خصوص در نزدیکی تکرار ۲۰۰ ام که مورد بحث ما نیز است. نمودارهای هر ۲ مدل دچار یک تغییر وضعیت نسبت به یکدیگر گشته و شرایط برای آن‌ها برعکس شده است. با توجه به مقادیر بدست آمده برای سرگشتگی بر روی پایگاه داده‌ی تست برای مدل پیشنهادی در این پژوهش در مقایسه با مدل RS در جدول ۵-۲، مشاهده می‌کنیم که در تمامی حالت‌ها رویکرد پیشنهادی مقدار کمتری را برای سرگشتگی محاسبه کرده است. لذا در تایید آنچه که گفتیم نتیجه گرفته می‌شود که راهکار پیشنهادی که با اضافه کردن یک لایه برای احساس نیز همراه است منجر به ساخت یک رویکرد احتمالی مناسب برای مدل‌سازی اسناد است که در مقایسه با مدل RS نتایج بهتری در بحث مدل‌سازی موضوع بدست می‌دهد.

۵-۱۰ مجسم‌سازی موضوع‌ها و ارزیابی دقت در محاسبه‌ی آن‌ها

در بخش ۵-۷ یک لغت‌نامه‌ی احساس به نام MPQA معرفی کردیم که شامل ۴۰۵۳ کلمه است که برچسب احساس برای آن‌ها مشخص شده است. در این بخش با استفاده از این لغت‌نامه‌ی احساس دقت موضوع‌های یاد گرفته شده توسط مدل را از نظر برچسب احساسی مورد ارزیابی قرار می‌دهیم.

با توجه به ساختار رویکرد پیشنهادی که در فصل ۵ توضیح داده شد، می‌دانیم که هر یک از واحدهای لایه‌ی Hidden به تمام واحدها چه در لایه‌ی Visible و چه در لایه‌ی Sentiment متصل هستند. هر واحد در لایه‌ی Sentiment برابر با یک برچسب احساسی و هر واحد در لایه‌ی Visible متناظر با یک کلمه است. از آنجا که در بحث مدل‌سازی موضوعی اسناد متنی، هر موضوع را به صورت یک توزیع احتمالی چند جمله‌ای بر روی تمام کلمات لغت‌نامه معرفی کردیم لذا می‌دانیم که هر واحد در لایه‌ی Hidden با یک وزن مشخص به تمام کلمات لغت‌نامه در لایه Visible متصل است. این وزن برای هر کلمه نشان دهنده‌ی مقدار اهمیت آن کلمه در

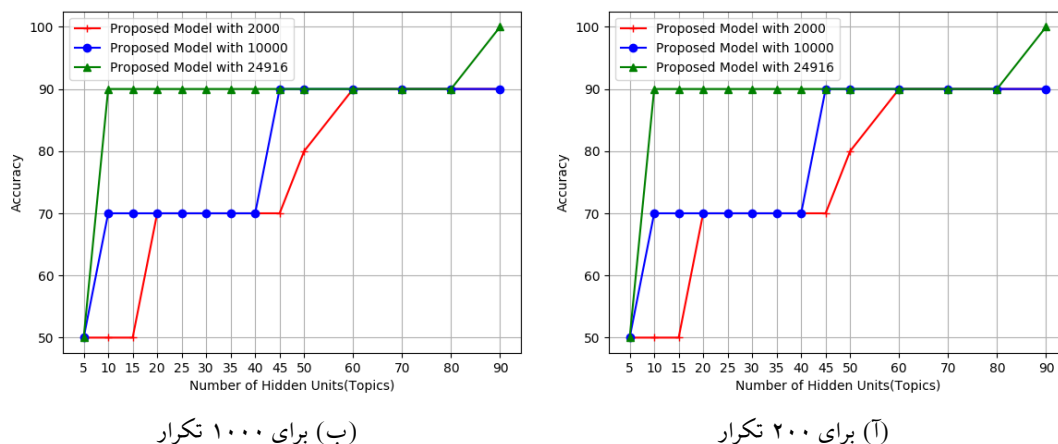
	Total Number	Numb of Positive Words	Num of Negative Words
NG(2000)	155	100	55
RCV(10000)	950	447	503
MR(24916)	3114	1242	1872

جدول ۵-۳: فراوانی‌های بدست آمده از مقایسه‌ی کلمات مشترک بین لغت‌نامه‌ی احساسی MPQA با سه لغت‌نامه واژگان

آن موضوع است. ایده‌ی ارزیابی مطرح شده در این بخش از آزمایش‌های انجام شده بر روی مدل‌های معروفی در زمینه‌ی مدل‌سازی موضوعی همچون DocNADE و LDA گرفته شده است. در مدل DocNADE که در بخش ۳-۳-۶ معرفی گردید و یک روش بر پایه‌ی شبکه عصبی است، برای مجسم‌سازی و نشان دادن موضوع‌های یاد گرفته شده توسط مدل از ماتریس وزن بین لایه‌ی *Hidden* و بردار کلمات استفاده می‌شود. به این صورت که برای هر موضوع وزن‌های تمام کلمات متصل به آن که در واقع تمام کلمات لغت‌نامه هستند مقایسه شده و بطور مثال ۱۰ کلمه‌ای که دارای بیشترین وزن برای هر موضوع هستند به عنوان کلمات متناسب با آن موضوع نشان داده می‌شوند. در مدل LDA نیز به همین صورت برای نمایش موضوع‌ها عمل می‌شود با این تفاوت که در روش LDA به جای ماتریس وزن یک ماتریس احتمال داریم و برای هر موضوع بطور مثال ۱۰ کلمه‌ای که بیشترین احتمال در آن موضوع را دارند نشان داده می‌شوند. در این بخش ما نیز از همین ایده کمک گرفته و موضوع‌ها را از نظر برچسب احساسی مورد ارزیابی قرار می‌دهیم.

ابتدا برای هر سه حالت مختلف از پایگاه داده تعداد کلمات مشترک با لغت‌نامه‌ی احساس MPQA را محاسبه کردیم. جدول ۵-۳ نتایج مربوط به این عمل را نشان می‌دهد. سپس به ازای هر ۳ حالت از پایگاه داده و ۲ تکرار مختلف برای مرحله‌ی آموزش و همچنین تعداد موضوع‌های مختلف مراحل زیر را به ترتیب انجام دادیم:

۱. محاسبه‌ی مجموع وزن‌های کلمات مثبت و منفی برای هر موضوع با استفاده از لغت‌نامه‌ی احساس و ماتریس وزن بین لایه‌ی *Visible* و *Hidden*.
۲. محاسبه‌ی تفاضل مقادیر حساب شده در مرحله ۱ برای هر موضوع و مرتب کردن مقادیر حاصل به صورت نزولی.
۳. انتخاب ۵ موضوع از ابتدای لیست مرتب (مثبت‌ترین موضوع‌ها) و تخصیص برچسب مثبت به آن‌ها، و ۵ موضوع از انتهای لیست مرتب (منفی‌ترین موضوع‌ها) و تخصیص برچسب منفی به آن‌ها.
۴. مقایسه‌ی برچسب تخصیص داده شده به هر موضوع با وزن‌های متناظر با آن موضوع در اتصال به لایه‌ی *Sentiment* و محاسبه‌ی دقت.



شکل ۵-۴: ارزیابی دقت در تخصیص احساس به موضوع‌ها به ۲۰۰ و ۱۰۰۰ مرحله آموزش

در مرحله‌ی ۱۴م منظور از مقایسه‌ی برچسب تخصیص داده شده به هر موضوع با وزن لایه‌ی *Sentiment* به این صورت است که اگر به یک موضوع در مرحله‌ی ۳ برچسب مثبت اختصاص داده شد، باید وزن متناظر با برچسب احساس مثبت برای آن موضوع در لایه‌ی *Sentiment* بیشتر از وزن منفی برای همان موضوع باشد و برعکس.

نمودارهای ۴-۵آ و ۴-۵ب نتایج حاصل از این ارزیابی را به ازای ۲۰۰ و ۱۰۰۰ مرحله آموزش نشان می‌دهند. همان‌طور که مشاهده می‌گردد برای هر ۳ حالت مختلف از پایگاه و تعداد موضوع‌های مختلف تفاوتی بین دقت محاسبه شده برای ۲۰۰ و ۱۰۰۰ مرحله‌ی آموزش وجود ندارد. اما با دقت در این نمودارها مشاهده می‌گردد که با بزرگ شدن اندازه لغت‌نامه دقت مدل در تخصیص برچسب احساسی به موضوع‌ها نیز افزایش می‌یابد.

مقایسه‌ی اطلاعات موجود در جدول ۳-۵ برای لغت‌نامه‌های مختلف با نمودارهای شکل ۴-۵ علت افزایش دقت به ازای افزایش اندازه لغت‌نامه را برای ما توجیه می‌کند. مشاهده می‌شود که با بزرگ شدن اندازه‌ی لغت‌نامه تعداد کلمه‌های مشترک بین آن و لغت‌نامه‌ی احساس نیز افزایش پیدا می‌کند و این امر سبب می‌گردد که در فرایند آموزش موضوع‌های مثبت و منفی بیشتر از یکدیگر تفکیک شده و در نتیجه دقت مدل در یادگیری و تخصیص برچسب احساس به موضوع‌ها افزایش پیدا می‌کند.

۱۱-۵ طبقه‌بندی احساسی اسناد

در این بخش نتایج حاصل از طبقه‌بندی احساس با استفاده از رویکرد پیشنهادی در این پژوهش را بر روی پایگاه داده‌ی MR ارزیابی و گزارش می‌کنیم. برای مقایسه‌ی نتایج بدست آمده در بحث طبقه‌بندی احساس با

استفاده از مدل پیشنهادی، از یک روش پایه که بر اساس شمارش تعداد کلمات است برای ارزیابی دقت در حالت‌های مختلف بهره می‌بریم. همچنین از نتایج بدست آماده برای طبقه‌بندی احساس با استفاده از چند روش معروف نظارت شده مانند ماشین بردار پشتیبان^۱ (SVM)، و دو شبکه عصبی (یک شبکه عصبی با مقادیر اولیه تصادفی و یک شبکه عصبی با مقادیر اولیه یادگرفته شده توسط رویکرد پیشنهادی) به منظور ارزیابی پارامترهای یاد گرفته شده توسط مدل استفاده می‌کنیم.

شبکه عصبی‌های استفاده شده در هر دو حالت (مقدار دهی تصادفی و مقدار دهی با پارامترهای یاد گرفته شده توسط مدل پیشنهادی) از دسته شبکه‌های MLP^۲ هستند. در لایه‌ی اول برای هر دو حالت تعداد نوروها برابر با تعداد موضوع‌ها و در لایه‌ی دوم تعداد نوروها برابر با تعداد احساس‌ها هستند. برای هردوی این شبکه‌ها از تابع خطای Cross Entropy استفاده شده است. همچنین در لایه‌ی اول این شبکه‌ها از تابع فعال‌ساز \tanh و در لایه‌ی دوم از تابع Softmax استفاده شده است.

برای محاسبه‌ی دقت در مدل پایه برای هر سند در پایگاه داده‌ی تست شروع به شمارش کلمات با قطبیت مشخص احساسی می‌کنیم. به بیان دیگر برای هر سند تعداد کلمات مثبت و تعداد کلمات منفی را با استفاده از لغت‌نامه‌ی احساس MPQA محاسبه می‌کنیم. بعد از محاسبه‌ی تعداد لغات مثبت و منفی برای هر سند اگر این مقدار برای کلمات مثبت در یک سند بیشتر از کلمات منفی بود به سند مورد نظر برچسب مثبت اختصاص دادیم و اگر تعداد کلمات منفی بیشتر بود آن سند را در دسته سندهای منفی دسته‌بندی می‌کنیم. برای طبقه بندی احساس به کمک رویکرد پیشنهادی در این پژوهش به این صورت عمل می‌کنیم که در ابتدا برای هر سند متنی با استفاده از

$$p(h_j = 1|\mathbf{V}) = \sigma \left(Db_j + \sum_{k=1}^K W_{kj} \hat{v}_k \right) \quad (2-5)$$

(این رابطه همان رابطه‌ی ۴-۲۳ است که برای سهولت در اینجا عیناً تکرار شده است) مقدار لایه‌ی مخفی متناظر با آن را بدست می‌آوریم. پس از محاسبه‌ی مقدار لایه‌ی پنهان با استفاده از رابطه‌ی ۵-۲ مرحله‌ی بعدی محاسبه‌ی لایه‌ی احساس متناظر با سند جاری با استفاده از

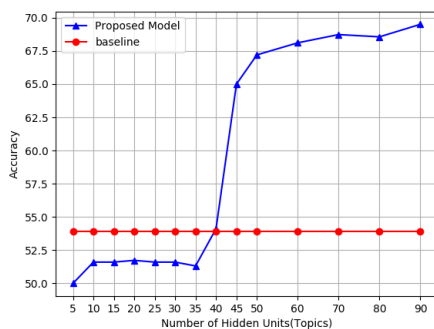
$$p(s_l = 1|\mathbf{h}) = \frac{\exp(c_l + \sum_{j=1}^H U_{lj} h_j)}{\sum_{l=1}^S \exp(c_l + \sum_{j=1}^H U_{lj} h_j)} \quad (3-5)$$

(این رابطه نیز همان رابطه‌ی ۴-۳۰ است) است.

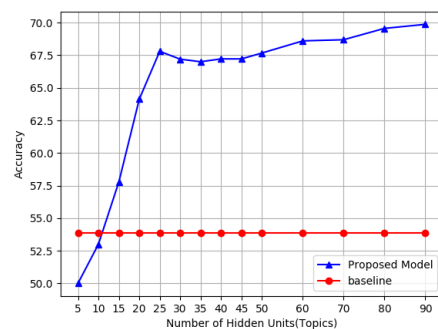
همان‌طور که در بخش ۴-۵ بیان گردید و در رابطه‌ی محاسبه لایه احساس (۴-۳۰) مشخص است، چون مقدار این لایه از یک تابع $Softmax$ بدست می‌آید لذا به فرم یک توزیع احتمالی است که مجموع درایه‌های

¹Support Vector Machine

²Multilayer Perceptron



(ب) با استفاده از لغت‌نامه با اندازه ۱۰۰۰۰



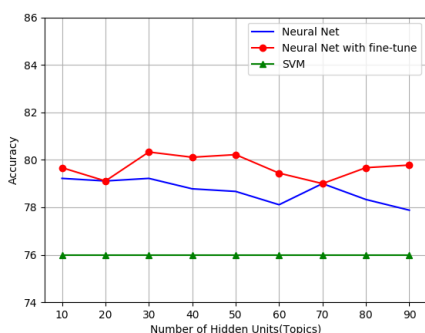
(آ) با استفاده از لغت‌نامه با اندازه ۲۰۰۰

شکل ۵-۵: طبقه‌بندی احساس در پایگاه داده‌ی MR با استفاده از مدل پیشنهادی و مدل پایه برای موضوع‌های مختلف

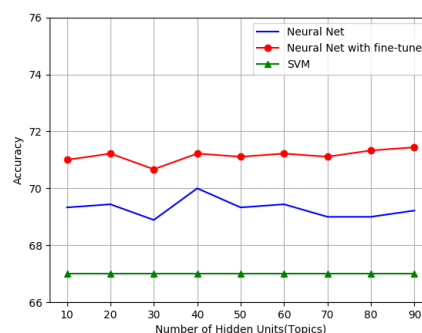
آن برابر با ۱ است. با بدست آوردن مقدار این لایه‌ی احساس سپس برای تخصیص برچسب به مقادیر این لایه نگاه می‌کنیم و مقدار متناظر با هراس احساس که بزرگتر بود برچسب آن سند را برابر با آن احساس انتخاب می‌کنیم. نتایج بدست آمده از طبقه‌بندی احساس با استفاده از رویکرد پیشنهادی و مدل پایه برای ۲ حالت مختلف از پایگاه داده در شکل ۵-۵ نشان داده شده است. برای محاسبه‌ی دقت در طبقه‌بندی احساس با استفاده از مدل پیشنهادی در این پژوهش برای هر ۲ حالت مختلف از پایگاه داده و تعداد موضوع‌های مختلف از مدل‌هایی که به ازای ۱۰۰۰ تکرار آموزش دیده‌اند استفاده شده است. همچنین برای مقدار دهی اولیه‌ی شبکه عصبی با پارامترهای یاد گرفته شده توسط رویکرد پیشنهادی، از مقادیر بدست آمده برای پارامترها (ماتریس وزن و بایاس) در این حالت (۱۰۰۰ چرخه آموزش) استفاده کردیم.

همانطور که در شکل‌های ۵-۵آ و ۵-۵ب مشاهده می‌گردد در هر ۲ حالت دقت طبقه‌بندی برای مدل پیشنهادی با افزایش تعداد موضوع‌ها رو به افزایش بوده است. همچنین در هر ۲ حالت دقت طبقه‌بندی با استفاده از مدل پیشنهادی با افزایش تعداد موضوع‌ها با اختلاف بسیار زیادی از دقت بدست آمده توسط مدل پایه بهتر است.

با مقایسه‌ی دقت بدست آمده برای طبقه‌بندی احساس به کمک رویکرد پیشنهادی و با استفاده از لغت‌نامه ۱۰۰۰۰ تایی (شکل ۵-۵ب) با نمودار شکل ۵-۴ب نتایج جالبی بدست می‌آید. در نمودار شکل ۵-۴ب دقت طبقه‌بندی احساس برای روش پیشنهادی در بازه‌ی تعداد موضوع‌های ۴۰ تا ۴۵ دچار یک جهش بزرگ و افزایش دقت با شیب زیادی می‌شود. با مقایسه این جهش و بررسی عمیق‌تر نمودار شکل ۵-۴ب مشاهده می‌شود که ارزیابی دقت در تخصیص برچسب احساس به موضوع‌ها برای لغت‌نامه ۱۰۰۰۰ تایی در بازه ۴۰ تا ۴۵ موضوع با یک افزایش با شیب بسیار زیاد همراه است. لذا نتیجه می‌شود که هرچه دقت در تخصیص برچسب احساس به موضوع‌ها بالاتر رود، یا به عبارت دیگر هرچه موضوع‌های متمایزتری از نظر احساسی داشته



۱۰۰۰۰ با استفاده از لغت‌نامه با اندازه (ب)



۲۰۰۰ با استفاده از لغت‌نامه با اندازه (آ)

شکل ۵-۶: طبقه‌بندی احساس در پایگاه داده‌ی MR با استفاده از مدل‌های شبکه عصبی با مقدار دهی اولیه برای وزن‌ها و بایاس‌ها، شبکه عصبی و SVM

باشیم دقت در بحث طبقه‌بندی احساس نیز بالاتر می‌رود.

شکل ۵-۶ دقت نتایج حاصل از طبقه‌بندی احساس با استفاده از ۳ مدل مختلف را نشان می‌دهد. با دقت در نمودارهای شکل ۵-۶ مشاهده می‌شود که در هر دو حالت مورد نظر برای پایگاه داده، دقت بدست آمده با استفاده از شبکه عصبی با مقدار دهی اولیه توسط پارامترهای یاد گرفته شده در روش پیشنهادی، از دقت بدست آمده توسط هر دو روش دیگر بهتر است.

در حالت استفاده از لغت‌نامه ۱۰۰۰۰ تایی همان‌طور که در شکل ۵-۶ مشخص است، تنها زمانی که تعداد موضوع‌ها برابر با ۲۰ و ۷۰ هستند دقت هر دو شبکه عصبی با هم برابر است. در سایر موارد شبکه با مقدار دهی اولیه‌ی پارامترها نسبت به هر ۳ مدل دیگر نتایج بهتری داشته است. به طور کلی نتیجه می‌شود که در فرآیند طبقه‌بندی احساس، شبکه‌ی عصبی که مقادیر آن توسط پارامترهای یاد گرفته شده توسط مدل پیشنهادی مقدار دهی اولیه می‌شوند از عملکرد بهتری نسبت به سایر مدل‌ها برخوردار است.

۵-۱۲ بازیابی اطلاعات

با توجه به اینکه رویکرد پیشنهادی در این پژوهش یک روش مولد برای مدل‌سازی همزمان احساس و موضوع است، لذا گام نخست برای ارزیابی این مدل در بحث بازیابی اطلاعات استفاده از پایگاه داده‌ای است که علاوه بر برچسب احساس برای اسناد، دارای برچسب موضوع برای هر سند نیز باشد. با توجه به عدم وجود یک چنین پایگاه داده‌ای، در این پژوهش ۲ پایگاه داده که همزمان شامل برچسب احساس و برچسب موضوع هستند ساخته شده است.

اولین پایگاه داده‌ی احساس-موضوع ساخته شده در این قسمت با تخصیص برچسب احساس به دیتاست ۲۰ NG ساخته می‌شود. در بخش ۵-۴ پایگاه داده‌ی ۲۰ NG را معرفی کردیم. همان‌طور که بیان شد ۲۰ NG

یک پایگاه داده‌ی معیار در بخش مدل‌سازی موضوعی است. اسناد موجود در ۲۰NG شامل ۲۰ گروه مختلف می‌شوند که هر یک از گروه‌ها در مورد یک موضوع خاص، مانند سیاسی، ورزشی، علمی و غیره هستند. برای اضافه کردن احساس به این مجموعه، مانند آنچه که در بخش ۵-۱۱ برای بدست آوردن دقت مدل پایه برای طبقه‌بندی احساس گفته شد، برای هر سند به شمارش تعداد کلمات با قطبیت مشخص احساسی با استفاده از لغت‌نامه‌ی احساس MPQA کردیم. سپس برای هر سند اگر تعداد کلمات مثبت بیشتر بود به آن سند برچسب مثبت اختصاص دادیم و برعکس. در فایل lib-svm که برای این پایگاه داده ساخته می‌شود، برای هر سند عدد اول نشان دهنده‌ی احساس (۱ مثبت، ۲ منفی) و عدد دوم (عددی از ۱ تا ۲۰) نشان دهنده‌ی موضوع مانند:

SentimentLabel TopicLabel <ID:Count> <ID:Count> <ID:Count> ...

برای آن سند است.

پایگاه داده‌ی دومی که برای ارزیابی رویکرد پیشنهادی در این پژوهش در بحث بازیابی اطلاعات ساخته می‌شود، از ترکیب چند دیتاست بدست می‌آید. پایگاه داده‌های MR و MDS در بخش‌های ۵-۳ و ۵-۵ به ترتیب معرفی شدند. هر کدام از این ۵ پایگاه داده (MDS شامل ۴ بخش مختلف با ۲۰۰۰ سند در هر بخش است) تنها شامل برچسب احساس هستند. می‌توان هر کدام از این مجموعه اسناد را به صورت یک موضوع خاص در نظر گرفت. به عبارت دیگر با کنار هم قرار دادن این پایگاه داده‌ها می‌توان یک پایگاه داده بزرگتر ایجاد کرد. این دیتاست جدید ساخته شده شامل ۱۰۰۰۰ سند است که ۵۰۰۰ تا از آن‌ها دارای برچسب مثبت و ۵۰۰۰ تای دیگر دارای برچسب منفی هستند. همچنین این پایگاه داده‌ی جدید شامل ۵ موضوع مختلف که بازیابی فیلم، کتاب، دی‌وی‌دی، وسایل آشپزخانه و وسایل الکترونیکی هستند، می‌شود. پس اتمام مرحله‌ی پیش‌پردازش هر کدام از این اسناد با استفاده از لغت‌نامه پایگاه داده‌ی ۲۰NG (۲۰۰۰ کلمه) به فایل libsvm تبدیل شدند. از ۱۰۰۰۰ سند نتیجه که با ترکیب این ۵ پایگاه داده بدست می‌آید، ۷۵۰۰ سند برای مجموعه آموزش با توزیع مساوی از نظر برچسب احساس (۳۷۵۰ سند مثبت و ۳۷۵۰ سند منفی) و موضوع (۱۵۰۰ سند از هر موضوع که ۷۵۰ تای آن مثبت و ۷۵۰ تای دیگر منفی هستند) انتخاب شدند و مابقی مجموعه‌ی تست را که شامل ۲۵۰۰ سند (۵۰۰ سند از هر موضوع که ۲۵۰ تای آن مثبت و ۲۵۰ تای آن منفی هستند) است، تشکیل می‌دهند. این پایگاه داده‌ی ساخته شده را به اختصار MRMDS نام‌گذاری می‌کنیم.

پس از ساخت پایگاه داده‌های مورد نیاز، به ارزیابی بازیابی اطلاعات برای رویکرد پیشنهادی در این پژوهش در مقایسه با مدل RS که در بخش ۳-۳-۵ معرفی شد می‌پردازیم. هدف از این ارزیابی مشاهده تاثیر در نظر گرفتن احساس برای بازیابی اطلاعات با استفاده از ساختار پیشنهادی در این پژوهش است. برای ارزیابی مورد

نظر از نمودار صحت در برابر فراخوانی^۱ استفاده می‌کنیم. این نمودار به عنوان معروف‌ترین معیار در بحث ارزیابی بازیابی اطلاعات و مقایسه‌ی روش‌های مختلف در این زمینه شناخته می‌شود. برای رسم این نمودار از مقادیر مختلف صحت و فراخوانی که توسط هر مدل بدست می‌آیند در برابر یکدیگر استفاده می‌شود. روابط

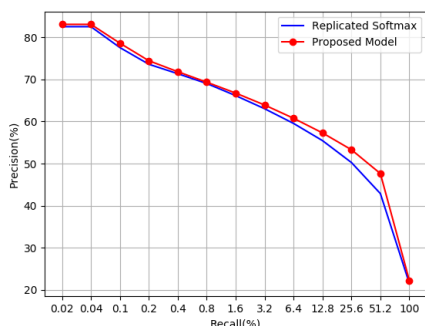
$$recall = \frac{TP}{C^+} \quad (۴-۵)$$

و

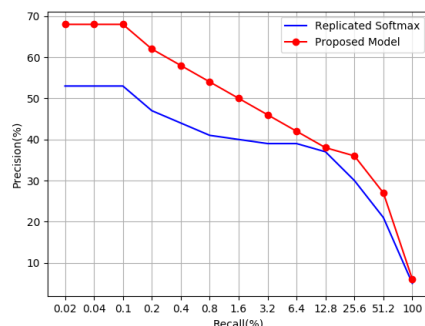
$$precision = \frac{TP}{R^+} \quad (۵-۵)$$

شیوه‌ی محاسبه‌ی مقادیر فراخوانی و صحت را نشان می‌دهند. منظور از TP در هر دو رابطه‌ی ۴-۵ و ۵-۵ مثبت‌های واقعی^۲ است که آن را برابر با تعداد اسنادی تعریف می‌کنیم که برچسب موضوعی برای آن‌ها همان برچسب مورد نظر ما است و مدل نیز آن‌ها را درست تشخیص داده است. C^+ در رابطه‌ی ۴-۵ نشان دهنده‌ی تعداد کل سندهای موجود با برچسب مورد نظر ما در پایگاه داده است و R^+ نشان دهنده‌ی تعداد اسنادی است که مدل آن‌ها را برای ما هماهنگ با برچسب مورد نظر ما تشخیص داده است.

نمودارهای شکل ۵-۵ نتایج حاصل از ارزیابی بازیابی اطلاعات برای رویکرد پیشنهادی در این پژوهش و همچنین مدل RS را نشان می‌دهند. همان‌طور که مشاهده می‌شود برای هر ۲ نمودار شکل ۵-۵ بخصوص نمودار ۵-۵آ روش پیشنهادی در این پژوهش عملکرد بهتری را در مقایسه با مدل RS در بحث بازیابی اطلاعات داشته است. برای محاسبه مقادیر صحت و فراخوانی و رسم نمودارهای شکل ۵-۵ به این صورت عمل شده است که، ابتدا بر روی هر کدام از پایگاه داده‌ها مدل پیشنهادی در این پژوهش بدون در نظر گرفتن برچسب موضوع و تنها با برچسب احساس و لایه‌ی مخفی با اندازه ۵۰، و همچنین روش RS بدون در نظر گرفتن برچسب‌های



(ب) پایگاه داده‌ی MRMDS



(آ) پایگاه داده‌ی ۲۰ News Groups

شکل ۵-۷: بازیابی اطلاعات با استفاده ۲ پایگاه داده‌ی ۲۰ News Groups و MRMDS برای رویکرد پیشنهادی و مدل RS

^۱Precision vs Recall

^۲True Positive

احساس و موضوع و لابه‌ی مخفی با اندازه ۵۰ به ازای ۵۰۰ تکرار آموزش داده شده‌اند. در مرحله‌ی بعدی برای تک‌تک سندهای مجموعه‌ی تست در هر پایگاه داده مقدار شباهت کسینوسی^۱ هر سند با تمام اسناد پایگاه داده‌ی آموزش محاسبه شده و مقادیر دقت و فراخوانی به دست آمده‌اند. در ادامه مقادیر بدست آمده برای صحت برای کل پایگاه داده‌ی تست میانگین گرفته می‌شوند و نمودارهای ۵-۵آ و ۵-۵ب رسم می‌شوند.

۵-۱۳ نتیجه‌گیری

در این فصل رویکرد پیشنهادی در این پژوهش که یک مدل مولد احتمالاتی نظارت شده برای مدل‌سازی مشترک موضوع و احساس در داده‌های متنی است مورد ارزیابی‌های مختلفی قرار گرفت. برای انجام این ارزیابی‌ها از چند پایگاه داده معروف در بحث مدل‌سازی موضوع و احساس استفاده گردید که هر کدام از آن‌ها در بخش‌های متفاوتی معرفی و به تفصیل شرح داده شدند. همچنین در این فصل یک لغت‌نامه‌ی احساس معرفی شد و با ساختار آن آشنا شدیم. در فصل بعدی نتیجه‌گیری نهایی از پژوهش و پیشنهادات آینده مطرح می‌گردند.

^۱ Cosine Similarity

فصل ششم

نتیجه‌گیری و پیشنهادات

۶-۱ نتیجه‌گیری

پیشرفت گسترده و چشمگیر در دنیای امروز که از آن به عنوان عصر ارتباطات و تکنولوژی یاد می‌کنیم منجر به تولید حجم بسیار زیادی از شکل‌های مختلف داده شده است. داده‌های تصویری، صوتی، متنی و غیره که هر روز مقدار آن‌ها بخصوص در بستر اینترنت رو به افزایش است. استفاده از این حجم عظیم از انواع مختلف داده که با سرعتی بسیار زیاد در حال افزایش نیز هستند نیاز به رویکردها و ساختارهایی دارد که توانایی تطبیق با این نرخ رشد را داشته باشند. مدل‌هایی که با پردازش اتوماتیک و خودکار این داده‌ها بتوانند از آن‌ها اطلاعات مفید و مورد نظر را استخراج کنند. بدون شک این حجم گسترده از شکل‌های مختلف داده حاوی اطلاعات سودمند بسیاری هستند که با استخراج آن‌ها می‌توان در کاربردهای گوناگون بهره‌های مناسبی از آن‌ها بدست آورد.

داده‌های متنی نیز در این میان از سهم بالایی برخوردار هستند. این شکل از داده هر روز در شبکه‌های اجتماعی، سایت‌های خرید و فروش، گروه‌های بحث و تبادل نظر، مجلات برخط و غیره در حال تولید و افزایش هستند. با تحلیل این داده‌ها، پردازش و استخراج اطلاعات از آن‌ها می‌توان به نتایج جالبی از قبیل مهم‌ترین موارد مطرح شده در شبکه‌های اجتماعی، احساس کلی کاربران و افراد جامعه در موارد خاص، بیشترین و مهم‌ترین موضوع‌های مطرح شده در مجلات و برخط و غیره بدست آورد. هدف ما در این پژوهش نیز در

همین راستا قرار داشت. ما به دنبال ساختاری مناسب و رویکردی جدید برای مدل‌سازی اسناد متنی واستخراج اتوماتیک اطلاعات، بخصوص اطلاعات مفهومی واحساس موجود در اسناد بودیم.

با توجه به مطالب بیان شده، هدف کلی در این پژوهش ساخت مدلی بر پایه‌ی شبکه‌های عصبی برای مدل‌سازی مشترک موضوع واحساس در داده‌های متنی است. در فصل سوم این پژوهش با روش‌های پیشین موجود در این زمینه آشنا شدیم. همان‌طور که گفته شد ساختارهای موجود در اکثر موارد تنها به مدل‌سازی موضوع و یا تشخیص احساس در پایگاه داده می‌پردازند. بررسی‌های انجام شده نشان داد که در زمینه‌ی مدل‌سازی مشترک موضوع واحساس تنها دو مدل ASUM (۱-۴-۳) و JST (۲-۴-۳) وجود دارند که با آن‌ها نیز به صورت کامل آشنا شدیم. گسترش شبکه‌های عصبی در سال‌های اخیر و استفاده‌ی فراوان از آن‌ها در بخش‌ها و زمینه‌های مختلف، همچنین عدم وجود ساختاری بر پایه‌ی شبکه‌های عصبی در زمینه‌ی مدل‌سازی مشترک احساس و موضوع به همراه کمبودها و کاستی‌های مدل‌های موجود در این زمینه دلایل اصلی نگارنده‌ی این پژوهش برای ساخت رویکردی جدید در این زمینه و انجام این پژوهش بوده است.

در این پژوهش یک رویکرد نظارت شده با استفاده از شبکه‌های عصبی برای مدل‌سازی مشترک موضوع واحساس در داده‌های متنی پیشنهاد شد. این ساختار که در دسته روش‌های احتمالاتی مولد دسته‌بندی می‌گردد با گسترش مدل معروف ماشین بلتزن محدود ایجاد می‌شود. در این رویکرد پیشنهادی یک لایه‌ی جدید با ماهیت توزیع احتمالاتی چندجمله‌ای به مدل اضافه شده و منجر به یادگیری ویژگی‌های بهتر و متمایز کننده‌تری برای هر سند در لایه‌ی مخفی می‌شود. همان‌طور که در فصل ۵ بیان گردید برای یادگیری و آموزش در این مدل از الگوریتم واگرایی مقابله استفاده می‌کنیم که یک روش تقریبی برای تخمین گرادیان است.

برای ارزیابی مدل پیشنهادی از پایگاه داده‌های بازبینی فیلم (MR)، ۲۰ گروه خبری (۲۰NG) و احساس چند دامنه (MDS) که همگی از پایگاه داده‌های شاخص در بحث مدل‌سازی موضوع و احساس در داده‌های متنی هستند، استفاده کردیم. با استفاده از یک معیار معروف برای ارزیابی مدل‌های مولد به نام سرگشتگی، رویکرد پیشنهادی در این پایان‌نامه را در فرآیند مدل‌سازی اسناد متنی ارزیابی کردیم. با توجه به نتایج بدست آمده در این بخش ادعا می‌کنیم که با در نظر گرفتن احساس موجود در اسناد و ایجاد ساختاری مانند آنچه که ما در این پژوهش انجام دادیم، یک مدل مولد بهتر برای مدل‌سازی اسناد ساخته می‌شود. همچنین با استفاده از پایگاه داده MR دقت رویکرد پیشنهادی را در زمینه‌های طبقه‌بندی احساس و ارزیابی موضوع‌های یاد گرفته شده مورد بررسی قرار دادیم و نتایج مربوط به آن را در بخش‌های ۵-۱۰ و ۵-۱۱ گزارش کردیم. در بخش ۵-۱۲ نیز رویکرد پیشنهادی در این پژوهش را در بحث ارزیابی اطلاعات در مقایسه با مدل RS مورد ارزیابی قرار دادیم. با انجام آزمایش بر روی ۲ پایگاه داده‌ی مختلف و مقایسه‌ی نتایج مشاهده گردید که روش پیشنهادی

نتایج بهتری را در بحث بازیابی اطلاعات بر روی داده‌های متنی دارد و از دقت بالاتری در این زمینه برخوردار است.

علی‌رغم اینکه در زمینه‌ی طبقه‌بندی احساس نتایج بدست آمده توسط رویکرد پیشنهادی در مقایسه با روش‌های موجود در بعضی حالت‌ها رقابتی است، اما مدل پیشنهادی در این پژوهش از نقاط قوت دیگری همچون نمایش موضوع‌های یاد گرفته شده و دقت بهتر در فرآیند مدل‌سازی به عنوان یک مدل مولد برخوردار است که آن را از دیگر روش‌های موجود در این زمینه متمایز می‌کند.

۶-۲ پیشنهادها

مدل پیشنهادی در این پژوهش (فصل ۵) از یک ساختار نظارت شده برای مدل‌سازی استفاده می‌کند. با توجه به اینکه حجم عظیمی از داده‌های موجود در بستر اینترنت داده‌های بدون برچسب هستند، لذا ساخت مدل‌های نیمه نظارتی، نظارت شده‌ی ضعیف و بدون نظارت به صورت قابل ملاحظه‌ای کاربردی و حائز اهمیت هستند. ساخت این دسته از مدل‌ها با گسترش رویکرد پیشنهادی از جمله مهم‌ترین کارهای پیش رو در این پایان‌نامه است.

پیشنهاد دیگری که در این قسمت برای کارهای آینده می‌تواند مطرح گردد استفاده از ایده‌ی ماشین بلترمن محدود شرطی و ارائه‌ی مدل‌های عمیق^۱ برای استخراج ویژگی‌های سطح بالاتر است. در ساختار پیشنهادی در این پایان‌نامه به ازای هر سند یک بردار ویژگی به نام لایه‌ی مخفی از آن استخراج می‌کنیم. پس از استخراج این بردار ویژگی از هر سند می‌توان خود این بردار را به عنوان یک داده‌ی ورودی برای یک شبکه عصبی RBM در نظر گرفت و مجدداً از آن ویژگی استخراج کرد. این ساختار منجر به یاد گرفتن ویژگی‌های متمایزتری برای هر داده‌ی ورودی می‌گردد.

رویکرد پیشنهادی در این پژوهش بر اساس کیسه‌ی کلمات عمل می‌کند. در استفاده از کیسه‌ی کلمات ساختار جملات و داده‌های ورودی از بین می‌روند و تنها تعداد تکرار کلمه‌ها مهم هستند. اما داده‌های متنی از سری داده‌های دنباله‌دار هستند و ترتیب در آن‌ها اهمیت دارد. لذا با در نظر گرفتن وابستگی کلمه‌ها در کنار یکدیگر و ترتیب آن‌ها می‌توان ویژگی‌های متمایز کننده‌تری از متن استخراج کرد. با توجه به اینکه در اکثر مواردی که بحث داده‌های دارای توالی مطرح می‌شود از شبکه‌های عصبی بازگشتی^۲ برای پیاده‌سازی استفاده می‌گردد، لذا پیشنهاد می‌کنیم در این حالت نیز از این شبکه‌ها استفاده شود.

^۱Deep Model

^۲Recurrent Neural Networks

مراجع

- [1] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.
- [2] H. Larochelle and I. Murray, "The neural autoregressive distribution estimator," in *The Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, vol. 15 of *JMLR: W&CP*, pp. 29–37, 2011.
- [3] G. E. Hinton and R. R. Salakhutdinov, "Replicated softmax: an undirected topic model," in *Advances in neural information processing systems*, pp. 1607–1614, 2009.
- [4] H. Larochelle and S. Lauly, "A neural autoregressive topic model," in *Advances in Neural Information Processing Systems*, pp. 2708–2716, 2012.
- [5] Y. Jo and A. H. Oh, "Aspect and sentiment unification model for online review analysis," in *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 815–824, ACM, 2011.
- [6] C. Lin, Y. He, R. Everson, and S. Ruger, "Weakly supervised joint sentiment-topic detection from text," *IEEE Transactions on Knowledge and Data engineering*, vol. 24, no. 6, pp. 1134–1145, 2012.
- [7] Z. Hai, G. Cong, K. Chang, P. Cheng, and C. Miao, "Analyzing sentiments in one go: A supervised joint topic modeling approach," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, pp. 1172–1185, June 2017.
- [8] Y. Zheng, Y.-J. Zhang, and H. Larochelle, "Topic modeling of multimodal data: an autoregressive approach," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1370–1377, 2014.
- [9] B. Pang, L. Lee, *et al.*, "Opinion mining and sentiment analysis," *Foundations and Trends® in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.
- [10] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural computation*, vol. 14, no. 8, pp. 1771–1800, 2002.
- [11] O. Woodford, "Notes on contrastive divergence," *Department of Engineering Science, University of Oxford, Tech. Rep.*, 2013.

- [12] M. A. Carreira-Perpinan and G. E. Hinton, "On contrastive divergence learning.," in *AISTATS*, vol. 10, pp. 33–40, Citeseer, 2005.
- [13] J. W. Mohr and P. Bogdanov, "Introduction—topic models: What they are and why they matter," *Poetics*, vol. 41, no. 6, pp. 545–569, 2013.
- [14] D. M. Blei, "Probabilistic topic models," *Communications of the ACM*, vol. 55, pp. 77–84, Apr. 2012.
- [15] D. M. Blei and M. I. Jordan, "Modeling annotated data," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pp. 127–134, ACM, 2003.
- [16] M. Steyvers and T. Griffiths, "Probabilistic topic models," *Handbook of latent semantic analysis*, vol. 427, no. 7, pp. 424–440, 2007.
- [17] P. Smolensky, "Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1," ch. Information Processing in Dynamical Systems: Foundations of Harmony Theory, pp. 194–281, Cambridge, MA, USA: MIT Press, 1986.
- [18] Y. Freund and D. Haussler, "Unsupervised learning of distributions on binary vectors using two layer networks," in *Advances in neural information processing systems*, pp. 912–919, 1992.
- [19] G. E. Hinton, *A Practical Guide to Training Restricted Boltzmann Machines*, pp. 599–619. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.
- [20] B. J. Frey, G. E. Hinton, and P. Dayan, "Does the wake-sleep algorithm produce good density estimators?," in *Advances in Neural Information Processing Systems 8* (D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, eds.), pp. 661–667, MIT Press, 1996.
- [21] Y. Bengio and S. Bengio, "Modeling high-dimensional discrete data with multi-layer neural networks," in *Advances in Neural Information Processing Systems*, pp. 400–406, 2000.
- [22] R. A. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1999.
- [23] G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*. New York, NY, USA: McGraw-Hill, Inc., 1986.
- [24] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *Journal of the American society for information science*, vol. 41, no. 6, p. 391, 1990.
- [25] T. Hofmann, "Probabilistic latent semantic indexing," in *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 50–57, ACM, 1999.
- [26] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pp. 79–86, Association for Computational Linguistics, 2002.
- [27] J. Blitzer, M. Dredze, F. Pereira, *et al.*, "Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification," in *ACL*, vol. 7, pp. 440–447, 2007.

Extension of Restricted Boltzman Machine for Joint Sentiment Topic Modeling in Text Data

Masoud Fatemi

m.fatemi@ec.iut.ac.ir

June , 2017

Department of Electrical and Computer Engineering
Isfahan University of Technology, Isfahan 84156-83111, Iran
Degree: M.Sc. Language: Farsi

Supervisor: Prof. Mehran Safayani (safayani@cc.iut.ac.ir)
Advisor: Prof. Abdolreza Mirzaei (mirzaei@cc.iut.ac.ir)

Abstract

Recently by the development of the internet and web, different types of social media such as web blogs became an immense source of text data. Through the processing of these data, not only is it possible to discover practical information about different topics, but it is also feasible to know about individual's opinions and gain a thorough understanding of the society sentiment. Therefore, having models which can automatically extract the subjective information from the documents would be efficient and helpful. Topic modeling procedures in addition data mining and also sentiment analysis are the most topics which considered in the natural language processing and text mining fields. Most of the existed models in these fields are based on the statistical methods and Bayesian networks so that there is no method for sentiment-topic modeling based on neural networks. Besides, the majority of existed frameworks have some constraint and difficulties such as computational complexity. This Thesis proposes a new structure for joint sentiment-topic modeling based on Restricted Boltzman Machine which is a type of neural networks. By altering the structure of Restricted Boltzman Machine plus appending a layer to it which is analogous to text data sentiment, feasibly we can propose a generative structure for joint sentiment topic modeling based on neural network. Proposed method is a supervised procedure which is trained by the Contrastive Divergence algorithm. The new attached layer in proposed model is a layer with the multinomial probability distribution quiddity which can be used in text data sentiment classification or any other supervised application. The proposed model is compared to existed models After implementing on document modeling as a generative model, sentiment classification and information retrieval the corresponding results have been reported thoroughly. It is observed that in the sentiment classification experiment the proposed model has a better accuracy on average of 11% in comparison with baseline model. Additionally, in the process of information retrieval on 20 news group's dataset, the proposed model by considering sentiment has an average of 9% enhanced function rather than the case in which sentiment is disregarded.

Key Words: Topic Model, Sentiment Analysis, Neural Networks, Restricted Boltzman Machine, Probabilistic Model, Contrastive Divergence Algorithm