

# مدل سازی مشترک موضوع و احساس در داده های متنی با استفاده از شبکه های عصبی

## چکیده

امروزه در حوزه ی هوش مصنوعی ما به دنبال الگوریتم ها و ساختارهایی هستیم که با دقت بالا یک رفتار انسانی و یا فرا انسانی را با بیشترین سرعت ممکن انجام دهند. با گسترش اینترنت و وب، انواع مختلف رسانه های اجتماعی نظیر وبلاگ ها و شبکه های اجتماعی در به یک منبع بسیار عظیم از انواع مختلف داده به ویژه داده های متنی تبدیل شده اند. با پردازش این داده ها می توان اطلاعات سودمند و مفیدی در مورد مباحث مختلف، نظر افراد و احساس کلی جامعه بدست آورد. از این جهت داشتن مدل هایی که کاملاً خودکار به تشخیص اطلاعات مفهومی و احساس در اسناد متنی بپردازند بسیار مفید است. روش های مدل سازی موضوع و استخراج اطلاعات مفهومی از داده های متنی و همچنین تشخیص احساس، همواره از مهمترین مباحث مطرح شده در زمینه ی پردازش زبان طبیعی، و کاوش داده های متنی است. بیشتر مدل هایی که در این زمینه وجود دارند بر پایه ی روش های آماری و شبکه های بیزی هستند به طوری که در زمینه ی مدل سازی موضوع-احساس با استفاده از شبکه های عصبی تا به امروز هیچ رویکردی وجود ندارد. همچنین بیشتر رویکردهای موجود دارای محدودیتهایی مانند پیچیدگی محاسباتی بالا هستند. در این مقاله یک ساختار جدید برای مدل سازی مشترک احساس-موضوع در داده های متنی بر پایه ی شبکه ی عصبی ماشین بلتزن محدود پیشنهاد می گردد. مدل پیشنهاد شده پس از پیاده سازی با مدل های موجود مقایسه گردید. مشاهده می شود رویکرد پیشنهادی در بحث ارزیابی به عنوان یک مدل مولد، طبقه بندی احساس و بازایی اطلاعات عملکرد بهتری نسبت به مدل های موجود دارد.

**کلمات کلیدی:** مدل سازی موضوع، آنالیز احساس، شبکه های عصبی، ماشین بلتزن محدود، مدل احتمالاتی، الگوریتم واگرایی مقابله

## ۱ مقدمه

امروزه در تمام مباحث مربوط به هوش مصنوعی ما به دنبال روش ها، الگوریتم ها 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 دسته بندی، استخراج اطلاعات مفهومی، آنالیز و برجسب گذاری داده ها و از جمله فعالیت هایی می باشند که امروزه ما انجام بسیاری از آن ها را به ماشین ها واگذار می کنیم.

در بین انواع مختلف داده، داده های متنی دارای سهم عظیمی از نظر حجم و مقدار هستند. به خصوص با گسترش اینترنت و وب در دهه ی اخیر با سرعتی بسیار زیاد، انواع مختلف رسانه های اجتماعی نظیر وبلاگ ها، شبکه های اجتماعی و گروه های بحث در اینترنت به یک منبع بسیار عظیم و قوی از انواع مختلف داده و اطلاعات به ویژه داده های متنی تبدیل شده اند. با پردازش این داده ها می توان اطلاعات سودمند و مفیدی در مورد مباحث مختلف، نقطه نظر افراد و احساس کلی جامعه بدست آورد. فعالیت های انجام گرفته در زمینه کاوش داده ها به خصوص کاوش داده های متنی و همچنین پردازش زبان طبیعی بیشتر از هر زمینه ی دیگری به تلاش برای درک و فهم این حجم عظیم از داده های متنی مربوط می شوند.

حجم عظیمی 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 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.

تمرکز ما در این مقاله و روش پیشنهادی پردازش بر روی داده‌های متنی است. در تقابل با داده‌های متنی، هدف ما پیدا کردن توزیع موضوع‌های مختلف موجود در مجموعه‌ی اسناد پایگاه داده و همچنین توزیع کلمات و احساس همراه با هر موضوع با استفاده شبکه‌ی عصبی است. فرآیند مورد نظر در داده‌های متنی تحت عنوان مدل‌سازی 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. 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 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 مقاله برای مدل‌سازی مشترک احساس و موضوع در داده‌های متنی بر پایه‌ی شبکه‌های عصبی گردید.

مدل‌سازی موضوع و استخراج اطلاعات مفهومی از داده‌های متنی و همچنین تشخیص احساس از مهمترین مباحث مطرح شده در زمینه‌ی پردازش زبان طبیعی، و کاوش داده‌های متنی هستند. رویکردهای موجود در این زمینه با اجرا بر روی یک پایگاه داده از اسناد متنی به تشخیص و مدل‌سازی موضوع‌ها، احساسات و مفاهیم همراه با هر سند متنی می‌پردازند. تشخیص احساس برای هر سند و هر موضوع در بحث بازیابی اطلاعات 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 پردازند می‌تواند بسیار مفید باشد. بیشتر کارهایی که در این زمینه وجود دارند بر پایه‌ی رویکردهای آماری و شبکه‌های بیزی هستند که از محدودیت‌هایی مانند پیچیدگی محاسباتی بالا رنج می‌برند. در بحث شبکه‌های عصبی بر خلاف مدل‌های آماری، روشی برای مدل کردن موضوع و احساس به صورت همزمان و مشترک وجود ندارد. در این مقاله نیز در همین راستا یک رویکرد نوین بر پایه‌ی شبکه‌های عصبی مصنوعی برای مدل‌سازی همزمان موضوع و احساس در یک مجموعه از داده‌های متنی پیشنهاد می‌گردد. رویکرد پیشنهاد شده در این مقاله یک مدل نظارت شده‌ی مولد احتمالی بر پایه‌ی شبکه‌ی عصبی ماشین بلتزن محدود است. برای آموزش در این مدل مانند سایر روش‌هایی که بر پایه‌ی ماشین بلتزن محدود هستند از الگوریتم یادگیری واگرایی مقابله استفاده می‌شود.

ساختار بخش‌های بعدی در مقاله به این صورت است: ابتدا در بخش دوم به مرور کارهای پیشین در زمینه‌ی تخمین توزیع‌های احتمالی در داده‌های ورودی، مدل‌سازی موضوع، تشخیص احساس و مدل‌سازی احساس-موضوع در داده‌های متنی می‌پردازیم. در بخش سوم کلیات نظری و تئوری مدل پیشنهادی بیان می‌شوند. در این فصل با معرفی یک مدل معروف به عنوان پایه مدل جدید تعریف و قسمت‌های مختلف آن شرح داده می‌شوند و روابط مورد نیاز برای هر قسمت تعریف می‌شوند. در بخش چهارم این مقاله مراحل شبیه‌سازی مدل پیشنهادی و نتایج حاصل از آزمایش‌های بدست آمده و مقایسه با دیگر مدل‌ها ارائه می‌گردد. در بخش پایانی، نتیجه‌گیری حاصل از این مقاله شرح داده خواهد شد. همچنین

راهکارهایی برای بهبود و توسعه مدل پیشنهادی ارائه خواهد شد.

## ۲ بررسی مدل‌های پیشین

در این بخش 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.

به طور کلی 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. در بررسی رویکردهای موجود از دید ساختاری، می‌توان آن‌ها را در دو گروه کلی طبقه‌بندی کرد. دسته‌ی اول مدل‌های گرافی و بیزی و دسته‌ی دوم مدل‌های بر پایه‌ی شبکه‌های عصبی. از نظر نحوه‌ی عملکرد مدل‌های پیشین را در سه کلاس مختلف قرار دارند. دسته‌ی اول روش‌هایی که تنها به مدل‌سازی موضوع می‌پردازند. دسته‌ی دوم روش‌هایی که به تشخیص احساس در داده‌های ورودی می‌پردازند. و در دسته‌ی سوم از نظر نحوه‌ی عملکرد روش‌هایی قرار دارند که به صورت همزمان به مدل‌سازی موضوع و احساس بر روی داده‌ی ورودی می‌پردازند. اگرچه باید توجه داشت که مدل‌های موجود در زمینه تشخیص احساس در دسته‌ی مدل‌های موضوعی قرار نمی‌گیرند و بیشتر شامل مدل‌هایی هستند که یک طبقه‌بندی دو حالتی (مثبت و منفی) یا سه حالتی (مثبت و منفی، مثبت و بی طرف) را انجام می‌دهند.

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 چندوجهی آن‌هایی هستند که شامل ترکیب چند حالت مختلف از داده‌ها می‌شوند، برای مثال ترکیب متن و تصویر و یا ترکیب تصویر و صدا.

از نقطه نظر سیر تکاملی می‌توان روش‌های موجود را در سه سطح: یک مدل‌های تخمین‌زننده‌ی توزیع، دو روش‌های مدل‌سازی موضوع و سه رویکردهای تشخیص همزمان موضوع و احساس 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.

در بحث مدل‌سازی موضوع، مدل‌های گرافی نسبت به روش‌های شبکه‌های عصبی از قدمت بیشتری برخوردار هستند. روش پایه‌ای که امروزه همچنان در مرورگرهای اینترنتی مورد استفاده قرار می‌گیرد، تکرار ترم-معکوس تکرار سند (tf-idf) نام دارد. این رویکرد در سال ۱۹۸۶ توسط Salton معرفی شد و در آن هر سند متنی به یک بردار اعداد حقیقی تبدیل می‌شود که شامل نسبت‌های تعداد تکرار کلمات مختلف است و از آن برای بازیابی اطلاعات استفاده می‌شود. برای غلبه بر محدودیت‌های tf-idf محققین حوزه‌ی IR چندین مدل کاهش بعد دیگر معرفی کردند که مهمترین آن‌ها مدل فهرست کردن معنایی نهفته (LSI) است که توسط Deerwester و همکاران در سال ۱۹۹۰ ارائه گردید. مدل LSI با استفاده از تجزیه مقدار منفرد بر روی ماتریس خروجی از مدل tf-idf یک زیرفضای خطی در فضای ویژگی‌های مدل

tf-idf شناسایی می‌کند. این روش منجر به کاهش قابل توجهی در 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 هستند، توانایی تشخیص بعضی از ویژگی‌های زبانی مانند مترادف و متضاد را دارند.

برای اثبات ادعاهای مطرح شده در مدل LSI و همچنین بررسی نقاط ضعف و قدرت این مدل، روش فهرست‌سازی معنایی نهفته‌ی احتمالاتی (pLSI) توسط Hofmann در سال ۱۹۹۹ معرفی شد. مدل pLSI یک مدل مولد احتمالاتی می‌باشد که از آن به عنوان یک مدل موضوعی نیز یاد می‌شود. در روش pLSI هر کلمه از یک موضوع خاص تولید می‌شود و کلمه‌های مختلف در داخل یک سند ممکن است از موضوع‌های مختلفی تولید شوند. مهم‌ترین مدلی که در دسته‌ی رویکردهایی گرافی وجود دارد مدل معروف تخصیص دیریکله‌ی پنهان (LDA) است که در سال ۲۰۰۳ توسط Blei و همکاران ارائه گردید و پس آن به عنوان پایه‌ی مدل‌سازی موضوعی در بخش مدل‌های گرافی قرار گرفت. 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.

مدل ماشین بلترن محدود که به اختصار آن را RBM می‌نامیم یک شبکه عصبی دولایه‌ی (یک لایه‌ی قابل مشاهده و یک لایه‌ی پنهان) بدون نظارت برای تخمین توزیع داده‌های ورودی باینری است. RBM یک مدل احتمالاتی مولد است که اولین بار در سال ۱۹۸۶ توسط Smolensky و پس از آن در سال ۲۰۰۲ به شکل دیگری توسط Hinton معرفی گردید. مدل شبکه‌ی عصبی خودکاهشی تخمین‌زننده‌ی توزیع (NADE) که از مدل RBM الهام گرفته شده است، یک روش احتمالاتی مولد بدون نظارت برای مدل‌سازی احتمال داده‌های گسسته است که در سال ۲۰۱۱ توسط Larochelle و همکاران ارائه شد. یکی از محدودیت‌های روش RBM مناسب نبودن این مدل برای تخمین احتمال مشترک در ابعاد بالا است. این محدودیت در مدل NADE بدلیل استفاده از ایده‌ی شبکه‌های بیزین کاملاً مشاهده‌پذیر برای محاسبه‌ی احتمال مرتفع گردیده است.

دسته‌ی دیگر مدل‌های موضوعی موجود از نظر ساختار آن‌هایی هستند که بر پایه‌ی شبکه‌های عصبی می‌باشند و اولین بار در سال ۲۰۰۹ توسط Hinton و Salakhutdinov تحت عنوان مدل سافتمکس تکرار شونده (RS) معرفی شدند. RS اولین روش مدل‌سازی موضوع بر پایه‌ی شبکه‌های عصبی و 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 موضوع‌ها مورد استفاده قرار بگیرد، چرا که اولاً کلمات باینری نیستند و دوماً در یک مجموعه از داده‌های متنی طول اسناد با یکدیگر متفاوت هستند. پس از مدل RS، شبکه‌ی عصبی خودکاهشی تخمین‌زننده‌ی توزیع سندی (DocNADE) یک روش بدون نظارت برای مدل‌سازی موضوع بر پایه‌ی شبکه‌های عصبی است که در سال ۲۰۱۲ توسط Larochelle و Lauly با ترکیب مدل‌های NADE و RS معرفی شد.

تمام مدل‌های بررسی شده تا کنون 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 ادامه دو رویکرد بر پایه‌ی شبکه‌های بیزی که در این دسته قرار دارند را معرفی می‌کنیم. مدل یکی‌سازی احساس-موضوع (ASUM) در سال ۲۰۱۱ برای تشخیص موضوع‌ها و احساس همراه با آن‌ها در بازیبنی‌های آنلاین توسط Jo و Oh معرفی شد. این

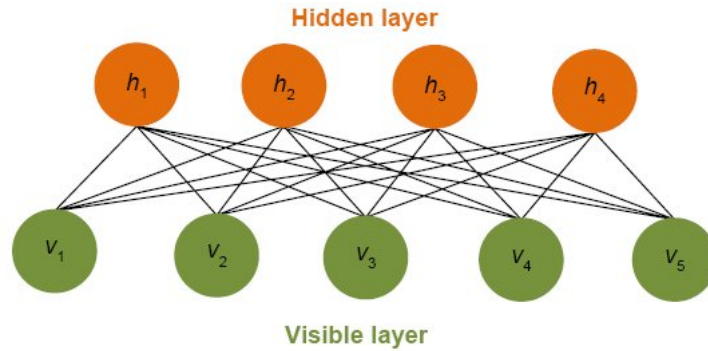
	Structure		Modeling Type			Data Input	
	Graphical	NN	DE	T	S/T	Unimodal	MultiModal
LSI	*			*		*	
pLSI	*			*		*	
LDA	*			*		*	
JST	*				*	*	
ASUM	*				*	*	
NADE		*	*			*	
RBM		*	*			*	
RS		*		*		*	
DocNADE		*		*		*	
SupDocNADE		*		*			*
Note: NN = Neural Network, DE = Distribution Estimator, T = Topic, S/T = Sentiment/Topic							

جدول ۱: دسته‌بندی مدل‌های پیشین از نظر ساختار، نحوه‌ی عملکرد و نوع داده‌ی ورودی.

مدل گسترش یافته‌ی مدل LDA است و در گروه مدل‌های احتمالاتی گرافیکی مولد قرار می‌گیرد. در مدل ASUM ما برای هر سند یک توزیع چندحمله‌ای احساسی و برای هر یک از احساس‌ها یک توزیع چندجمله‌ای موضوعی داریم و فرض بر این است که هر جمله در داخل هر سند دارای یک برچسب احساس و یک موضوع است. پس از روش ASUM، مدل نظارت‌شده‌ی ضعیف تشخیص مشترک احساس-موضوع (JST) در سال ۲۰۱۲ توسط Lin و همکاران معرفی شد. مدل JST یک مدل احتمالاتی مولد گرافیکی و گسترش یافته‌ی مدل LDA است که علاوه بر تشخیص موضوع به تشخیص احساس از داده‌های متنی نیز می‌پردازد. خاصیت نظارت‌شده‌ی ضعیف باعث می‌شود که در مقایسه با سایر مدل‌ها، JST به راحتی قابل انتقال به یک دامنه‌ی دیگر بدون کاهش محسوس در کارایی که در سایر مدل‌ها این اتفاق رخ می‌دهد باشد. مدل‌های چندحالتی (Multimodal) دسته‌ای از مدل‌های موضوعی هستند که داده‌ی ورودی در آن‌ها ترکیبی از چند حالت مختلف داده است. مدل نظارت‌شده‌ی شبکه‌ی عصبی خودکاهشی تخمین‌زننده‌ی توزیع سندی (SupDocNADE) یک رویکرد چندحالتی است که در سال ۲۰۱۴ توسط Zheng و همکاران معرفی شد. این مدل گسترش یافته‌ی مدل DocNADE است. 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 گرفته و در کاربرد مورد نظر از آن استفاده می‌کند. در جدول ۱ به صورت خلاصه ویژگی‌های مدل‌های معرفی شده نشان داده شده و این مدل‌ها در سه کلاس مختلف دسته‌بندی گردیده‌اند.

### ۳ مدل‌سازی مشترک موضوع و احساس با شبکه‌های عصبی

مدل پایه برای روش پیشنهادی در این مقاله شبکه عصبی RBM است که در شکل ۱ نشان داده شده است. RBM یک مدل بدون نظارت برای داده‌های باینری است که در دسته‌ی مدل‌های مولد احتمالاتی قرار می‌گیرد. در این مدل با بیشینه



شکل ۱: ماشین بلتزمن محدود RBM

کردن یک تابع انرژی، یا کمینه کردن مقدار منفی آن که به صورت رابطه ۱ تعریف می‌شود، توزیع‌های احتمالی موجود در داده‌های ورودی یاد گرفته می‌شود و از داده‌های ورودی ویژگی استخراج می‌گردد. در رابطه‌ی ۱،  $\theta = \{W, \mathbf{a}, \mathbf{b}\}$  مجموعه‌ی پارامترهای مدل است.  $W_{D \times H}$  ماتریس وزن بین لایه‌ی ورودی و لایه‌ی پنهان است، که در آن  $D$  سایز بردار ورودی و  $H$  سایز لایه‌ی پنهان هستند.  $\mathbf{a}$  بردار بایاس لایه‌ی ورودی با سایز  $D$  و  $\mathbf{b}$  بردار بایاس لایه‌ی پنهان با سایز  $H$  است. در ساختار RBM توزیع‌های شرطی برای لایه‌های قابل مشاهده و پنهان به شکل روابط ۲ و ۳ هستند.

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

$$p(\mathbf{h}|\mathbf{v}) = \text{sigmoid}(\mathbf{v}W + \mathbf{b}) \quad (۲)$$

$$p(\mathbf{v}|\mathbf{h}) = \text{sigmoid}(W\mathbf{h} + \mathbf{a}) \quad (۳)$$

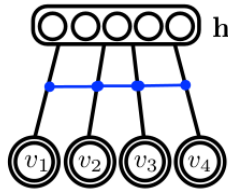
در مدل RBM احتمال هر ترکیب  $(\mathbf{v}, \mathbf{h})$  از رابطه‌ی ۴ بدست می‌آید که در آن  $Z(\theta)$  تابع قسمت‌بندی است که مقدار آن با استفاده از رابطه‌ی ۵ محاسبه می‌شود و تضمین می‌کند که مقدار بدست آمده برای هر ترکیب  $(\mathbf{v}, \mathbf{h})$  در رابطه‌ی ۴ یک مقدار صحیح احتمالی (بین ۰ و ۱) است. در این مدل احتمال هر بردار ورودی از رابطه‌ی ۶ بدست می‌آید.

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

$$Z = \sum_{\mathbf{v}, \mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h})} \quad (۵)$$

$$p(\mathbf{v}) = \sum_h \frac{1}{Z(\theta)} e^{-E(\mathbf{v}, \mathbf{h})} \quad (۶)$$

محدود بودن به حالت باینری برای داده‌های ورودی و همچنین طول ثابت برای آن‌ها دو محدودیت اساسی در مدل RBM استاندارد هستند. فرض کنید در مساله‌ای که با آن سرو کار داریم هر داده‌ی ورودی دارای  $D$  ویژگی است که هر یک از این



شکل ۲: مدل Replicated Softmax

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 سایز  $K \times D$  در نظر گرفته می‌شود که همان‌طور که بیان شد  $D$  طول بردار ورودی یا همان تعداد ویژگی‌های مساله و  $K$  ماکسیمم مقداری است که هر ویژگی می‌تواند داشته باشد. در این صورت تابع انرژی برای حالت  $\{\mathbf{V}, \mathbf{h}\}$  به شکل رابطه‌ی ۵ تعریف می‌شود. همچنین رابطه‌ی ۳ به شکل رابطه‌ی ۸ تبدیل می‌شود. دلیل استفاده از تابع Softmax در رابطه‌ی ۸ به جای تابع سیگموئید که در مدل RBM استاندارد از آن استفاده می‌شود 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، لذا برای هر ستون در این ماتریس از تابع Softmax استفاده می‌گردد و یک توزیع احتمال چند جمله‌ای بدست می‌آید. سپس با تولید نمونه از این توزیع چند جمله‌ای مقدار آن ویژگی تعیین می‌گردد.

$$E(\mathbf{V}, \mathbf{h}) = - \sum_{i=1}^D \sum_{j=1}^H \sum_{k=1}^K W_{ijk} h_j v_{ik} - \sum_{i=1}^D \sum_{k=1}^K v_{ik} a_{ik} - \sum_{j=1}^H h_j b_j \quad (7)$$

$$p(v_{ik} = 1 | \mathbf{h}) = \frac{\exp(a_{ik} + \sum_{j=1}^H h_j W_{ijk})}{\sum_{k=1}^K \exp(a_{ik} + \sum_{j=1}^H h_j W_{ijk})} \quad (8)$$

برای مدل‌سازی موضوع پیش از آموزش مدل ابتدا یک لغت‌نامه از تمام کلمات متمایز در مجموعه اسناد ساخته می‌شود. حال در بردار ورودی مقدار هر ویژگی برابر با اندیس یکی از کلمات دیکشنری است. به بیان دیگر هر سند ورودی پس از انجام پیش پردازش‌های لازم به یک دنباله از کلمات تبدیل می‌شود که هر کدام از این کلمات برابر با یکی از کلمات دیکشنری هستند. به این ترتیب در ماتریس ورودی به مدل که یک ماتریس به اندازه‌ی  $K \times D$  است،  $K$  برابر با سایز دیکشنری و  $D$  نشان دهنده‌ی طول سند متنی است. در این حالت برای هر ستون مقدار سطر متناظر با اندیس آن کلمه در دیکشنری برابر با ۱ می‌شود و دیگر درایه‌های آن ستون همچنان صفر باقی می‌مانند. در این مدل که Hinton و Salakhutdinov آن را RS نامیدند و در شکل ۲ نشان داده شده است، برای مدل کردن داده‌های متنی برای هر سند یک شبکه‌ی 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 رابطه‌ی ۹ است که در آن  $\hat{v}^k = \sum_{i=1}^D v_{ik}$  است. به عبارت دیگر  $\hat{v}$  برداری است با طول  $K$ ، که همان سایز دیکشنری است و از محاسبه‌ی حاصل جمع سطرهای ماتریس باینری ورودی بدست می‌آید. در این حالت روابط شرطی محاسبه‌ی لایه‌ی قابل مشاهده و لایه‌ی پنهان به شکل روابط ۱۰

و ۱۱ هستند. همانطور که مشاهده می‌شود در روابط ۹ و ۱۱ ترم بایاس برای لایه‌ی پنهان با سائز سند جاری نیز متناسب است. وجود این تناسب در پیاده‌سازی‌های تجربی و هنگامی که اسناد با طول‌های متفاوت در مجموعه اسناد وجود دارد بسیار حیاتی است.

$$E(\mathbf{v}, \mathbf{h}) = - \sum_{j=1}^H \sum_{k=1}^K W_{jk} h_j \hat{v}_k - \sum_{k=1}^K v_k a_k - D \sum_{j=1}^H h_j b_j \quad (9)$$

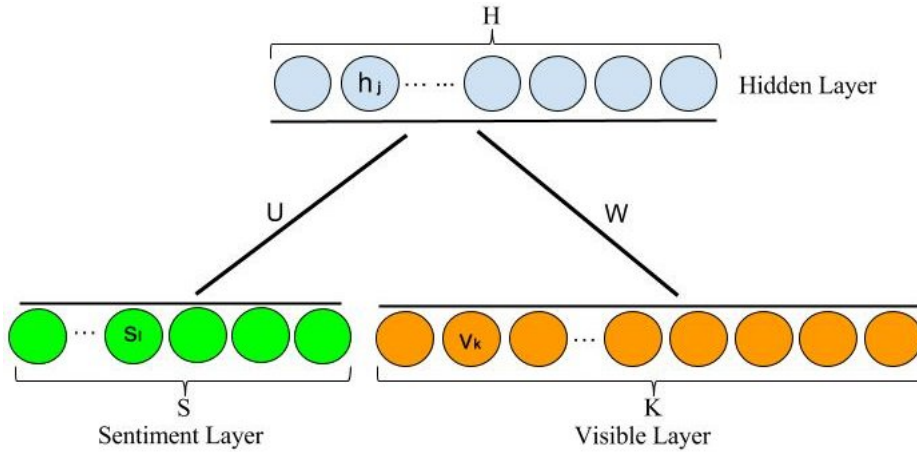
$$p(v_i = w | \mathbf{h}) = \frac{\exp(a_w + \sum_{j=1}^H h_j W_{wj})}{\sum_{k=1}^K \exp(a_k + \sum_{j=1}^H h_j W_{kj})} \quad (10)$$

$$p(h_j = 1 | \mathbf{v}) = \sigma \left( D b_j + \sum_{k=1}^K W_{kj} \hat{v}_k \right) \quad (11)$$

مدل‌ها و حالت‌های معرفی شده تاکنون رویکردهایی هستند که پایه‌ی و اساس ساختار پیشنهادی در این مقاله هستند و با گسترش آن‌ها به شکلی که در ادامه بیان می‌شود مدل پیشنهادی در این مقاله بدست می‌آید. رویکرد معرفی شده در این مقاله یک مدل مولد احتمالاتی نظارت شده بر پایه‌ی شبکه‌ی عصبی برای مدل‌سازی موضوع و احساس در داده‌های متنی است که در شکل ۳ 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 برچسب هر سند یا تعداد کلاس‌های موجود 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 قابل مشاهده یک بردار با طولی ثابت و به اندازه‌ی سائز دیکشنری یا همان تعداد کلمات متمایز در متن است که در آن تعداد تکرار کلمات مشخص شده است. در نظر گرفتن یک ماتریس ۲ بعدی برای هر سند ورودی به این معنی است که جایگاه هر کلمه در متن دارای اهمیت 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 رفتار می‌کنند که در آن ترتیب کلمات در نظر گرفته نمی‌شود.

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 ۳ نشان داده شده است طول بردار ورودی برابر با سائز دیکشنری یا همان تعداد کلمات متمایز در مجموعه سند در نظر گرفته می‌شود که درایه‌های آن تعداد تکرار هر کلمه از دیکشنری 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 دیکشنری یک وزن و یک بایاس تعریف می‌گردد و این کلمه در هر جای سند ورودی قرار داشته باشد وزنش تغییری نخواهد کرد و در فرآیند آموزش تنها یک وزن و یک بایاس برای هر کلمه یاد گرفته می‌شود. با توجه به ساختار مدل که در شکل ۳ نشان داده شده است، برای هر سند متنی یک بردار باینری که نشان دهنده احساس سند جاری است به عنوان ورودی به شبکه وارد می‌شود، و توزیع‌های موجود بر روی کلمات مختلف در هر موضوع و همچنین احساس مرتبط با آن‌ها توسط مدل در لایه‌ی پنهان استخراج می‌شوند.





شکل ۳: مدل پیشنهادی مولد احتمالی احساس/موضوع

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 ۱۲، با استفاده از فرمول ۱۳ 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 به شکل رابطه‌ی ۱۴ استفاده می‌شود. در رابطه ۱۲،  $\theta = \{W, U, \mathbf{a}, \mathbf{b}, \mathbf{c}\}$  مجموعه پارامترهای مدل است که در آن  $W_{K \times H}$  ماتریس وزن بین بردار قابل مشاهده و لایه‌ی پنهان،  $U_{S \times H}$  ماتریس وزن بین لایه‌ی احساس و لایه‌ی پنهان و  $\mathbf{a}$ ،  $\mathbf{b}$  و  $\mathbf{c}$  به ترتیب بردارهای بایاس لایه‌ی قابل مشاهده، پنهان و احساس هستند. لازم به ذکر است که  $H$  و  $K$  مانند آنچه که بیش از این ذکر کردیم به ترتیب سائز دیکشنری و طول لایه‌ی پنهان هستند و  $S$  به عنوان تعداد احساس موجود یا سائز بردار احساس تعریف می‌شود.

$$E(\mathbf{V}, \mathbf{s}, \mathbf{h}) = - \sum_{j=1}^H \sum_{k=1}^K W_{kj} h_j \hat{v}_k - \sum_{j=1}^H \sum_{l=1}^S U_{lj} h_j s_l - \sum_{k=1}^K v_k a_k - \sum_{l=1}^S s_l c_l - D \sum_{j=1}^H h_j b_j \quad (12)$$

$$p(\mathbf{v}, \mathbf{s}, \mathbf{h}) = \frac{1}{Z} e^{-E(\mathbf{v}, \mathbf{s}, \mathbf{h})} \Rightarrow p(\mathbf{v}, \mathbf{s}) = \frac{1}{Z} \sum_{\mathbf{h}} e^{-E(\mathbf{v}, \mathbf{s}, \mathbf{h})}, Z = \sum_{\mathbf{v}} \sum_{\mathbf{s}} \sum_{\mathbf{h}} e^{-E(\mathbf{v}, \mathbf{s}, \mathbf{h})} \quad (13)$$

$$\Delta \theta = \alpha (E_{P_{data}}[\theta] - E_{P_{model}}[\theta]) \Rightarrow \theta_{t+1} = \theta_t + \Delta \theta \quad (14)$$

در مدل پیشنهادی مقادیر هر یک از لایه‌های قابل مشاهده، احساس و پنهان به کمک روابط ۱۵ تا ۱۷ محاسبه می‌شوند. در اینجا چون مقدار لایه‌ی پنهان به هر دو مقدار لایه‌ی قابل مشاهده و احساس وابسته است، لذا مشاهده می‌شود که در رابطه‌ی ۱۷ برای مقدار لایه‌ی پنهان از یک توزیع شرطی که وابسته به هر دو مقدار لایه‌های قابل مشاهده و احساس

است نمونه گرفته می‌شود. اما با توجه به اینکه با داشتن مقدار لایه‌ی پنهان، بردارهای قابل مشاهده و احساس از یکدیگر مستقل شرطی هستند لذا در روابط ۱۵ و ۱۶ مقدار این دو بردار از یک توزیع شرطی که تنها به مقدار بردار پنهان وابسته است نمونه گرفته می‌شوند.

$$p(v_i = w|\mathbf{h}) = \frac{\exp(a_w + \sum_{j=1}^H W_{wj}h_j)}{\sum_{k=1}^K \exp(a_w + \sum_{j=1}^H W_{wj}h_j)} \quad (15)$$

$$p(s_l = 1|\mathbf{h}) = \frac{\exp(c_l + \sum_{j=1}^H U_{lj}h_j)}{\sum_{l=1}^S \exp(c_l + \sum_{j=1}^H U_{lj}h_j)} \quad (16)$$

$$p(h_j = 1|\mathbf{v}, \mathbf{s}) = \sigma \left( Db_j + \sum_{k=1}^K W_{kj}\hat{v}_k + \sum_{l=1}^S U_{lj}s_l \right) \quad (17)$$

با توجه به خصوصیات بیان شده برای تابع Softmax، لذا همان‌طور که مشاهده می‌شود، در روابط ۱۵ و ۱۶ برای محاسبه‌ی مقادیر لایه‌های قابل مشاهده و احساس از یک تابع Softmax استفاده می‌گردد. در فرآیند آموزش با استفاده از الگوریتم CD برای بدست آوردن مقدار بازسازی شده از لایه‌ی قابل مشاهده مشروط به بردار پنهان از رابطه‌ی ۱۵ که به صورت Softmax است، استفاده می‌شود. در واقع دلیل اینکه این رابطه و رابطه‌ی ۱۶ برای لایه‌ی احساس به فرم تابع Softmax هستند همین امر می‌باشد، که پس از محاسبه‌ی مقادیر این لایه‌ها مشروط به بردار پنهان نیاز به تولید نمونه و نمونه‌برداری از این مقادیر بدست آماده 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 آزمایش‌ها و ارزیابی مدل

## ۵ نتیجه‌گیری