بسم الله الرحمن الرحیم

دانشکده مهندسی کامپوتر و فناوری اطلاعات

رشته مهندسی کامپیوتر، گرایش هوش مصنوعی

پایان‌نامه کارشناسی ارشد

سیستم های پرسش- پاسخ و ابزار یادگیری عمیق

نگارنده: نوید سلیمانی

استاد راهنما

دکتر مرتضی زاهدی

استاد مشاور

دکتر حمید حسن‌پور

تیر ۱۳۹۸

با سپاس از سه وجود مقدس:
آنان که موهایشان سپید شد تا ما روی سفید شویم... .

پدر

مادر

معلم

سپاس‌گزاری...

از معلم ارجمندم جناب آقای مرتضی زاهدی، تقدیر و تشکر می کنم، چرا که همواره مشوق و راهنمای من در پیشبرد و پیشرفت بوده‌اند.

نوید سلیمانی
تیر ۱۳۹۸

تعهد نامه

اینجانب **نوید سلیمانی** دانشجوی کارشناسی ارشد رشته **مهندسی کامپیوتر مهندسی کامپوتر و فناوری اطلاعات** دانشگاه شاهرود، نویسنده پایان‌نامه با عنوان **سیستم های پرسش-پاسخ و ابزار یادگیری عمیق** ، تحت راهنمایی **مرتضی زاهدی** متعهد می‌شوم:

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های دیگر پژوهش‌گران، به مرجع مورد استفاده استناد شده است.
- مطالب این پایان‌نامه، تا کنون توسط خود، یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ‌جا ارایه نشده است.
- حقوق معنوی این اثر، به دانشگاه صنعتی شاهرود تعلق دارد، و مقالات مستخرج با نام " دانشگاه صنعتی شاهرود " یا " Shahrood University of Technology " به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به‌دست آوردن نتایج اصلی پایان‌نامه تاثیرگذار بوده‌اند، در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافت‌های آنها) استفاده شده است، ضوابط و اصول اخلاقی رعایت شده است.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته (یا استفاده شده است)، اصل رازداری و اصول اخلاق انسانی رعایت شده است.

نوید سلیمانی
تیر ۱۳۹۸

مالکیت نتایج و حق نشر

- تمام حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته شده) متعلق به دانشگاه صنعتی شاهرود می‌باشد. این مطلب باید به نحو مقتضی، در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در این پایان‌نامه بدون ذکر منبع مجاز نمی‌باشد.

چکیده

در طی چند سال گذشته محققان حوزه‌ی پردازش‌های زبان‌های طبیعی به کمک مدل‌های یادگیری عمیق پیشرفت فراوانی را به دست آورده‌اند و سیستم‌های پرسش‌ـپاسخ به عنوان زیر مجموعه‌ای از علوم پردازش زبان‌های طبیعی از این جریان مستثنا نبوده است. در این پایان نامه قصد داریم یک مدل مبتنی بر یادگیری عمیق برای پاسخ خودکار به پرسش ارائه دهیم. با توجه با آن می‌توانیم که سیستم پرسش‌ـپاسخ را به سه جز: طبقه بندی پرسش‌ها و پاسخ‌ها، بازیابی اطلاعات و انتخاب پاسخ، تقسیم کنیم. ابتدا پرسش‌ها و پاسخ‌ها را با استفاده از یک مدل‌سازی احتمالاتی عصبی طبقه بندی و به بردار تبدیل می‌کنیم. سپس شبکه عصبی شباهت عمیق را برای پیدا کردن میزان شباهت بین پرسش و پاسخ آموزش می‌دهیم و در انتها بهترین پاسخ را انتخاب و ارائه می‌دهیم.

سیستم پیشنهادی ما، با پیشنهاد روشی مبتنی بر مدل‌سازی احتمالاتی عصبی، عبارات را به گونه‌ای جاسازی و به بردار تبدیل می‌کند تا بردار هر عبارت بسیار شبیه به بردار کلمه یا کلمات کلیدی آن شود. در نتیجه می‌توان احتمال شباهت پرسش با پاسخ‌های مربوط به آن را بالا برد و در انتها شبکه‌ی عصبی سیام‌ـ با اعمال تغییراتی‌ـ به گونه‌ای آموزش می‌بیند تا بتواند پاسخ مناسبی را برای پرسش مطرح شده، انتخاب کند. در این مدل ما توانسته‌ایم خطا را حدود ۲۵ درصد نسبت به روش‌های مشابه اخیر کاهش دهیم.

کلمات کلیدی:

هوش مصنوعی، پردازش زبان‌های طبیعی، سیستم‌های پرسش‌ـپاسخ، ابزار یادگیری عمیق، مدل‌سازی احتمالاتی عصبی، شبکه‌های عصبی شباهت عمیق.

فهرست مطالب

فهرست تصاویر ف

فهرست جداول ق

۱ مقدمات ۱

۱.۱ مقدمه ۱

۲.۱ یک زمینه ی تحقیقات طولانی ۲

۱.۲.۱ آغاز مسأله‌ی پرسش ـ پاسخ ۲

۲.۲.۱ برخی از مسائل قابل بحث در زمینه ی پرسش ـ پاسخ ۳

۳.۱ بیان مسأله ۴

۱.۳.۱ سیستم‌های پرسش ـ پاسخ خودکار ۴

۲.۳.۱ پیدا کردن شباهت بین جفت پرسش ـ پاسخ ۴

۳.۳.۱ استفاده از یادگیری عمیق ۵

۴.۱ اهمیت و ضرورت تحقیق ۵

۵.۱ اهداف ۵

۶.۱ تعریف‌های اولیه ۵

۱.۶.۱ پردازش زبان‌های طبیعی ۵

۲.۶.۱ سیستم‌های بازیابی اطلاعات ۵

۳.۶.۱ مسأله‌ی پرسش ـ پاسخ ۶

۴.۶.۱ شبکه‌های عصبی مصنوعی ۶

۵.۶.۱ یادگیری عمیق ۶

۶.۶.۱ مدل‌سازی احتمالاتی عصبی ۶

۷.۶.۱ شبکه‌های عصبی مصنوعی شباهت عمیق ۷

۷.۱ جمع بندی ۷

۲ ادبیات و پیشینه تحقیق ۹
۱.۲ مقدمه ۹
۲.۲ مروری بر تحقیقات پیشین ۱۰
۱.۲.۲ پردازش زبان‌های طبیعی ۱۰
۲.۲.۲ شبکه‌های عصبی مصنوعی و پردازش زبان‌های طبیعی ۱۱
۳.۲.۲ شبکه‌های عصبی مصنوعی پیش‌خور ۱۳
۴.۲.۲ شبکه‌های عصبی مصنوعی بازگشتی ۱۴
۵.۲.۲ شبکه‌های عصبی مصنوعی حافظه طولانی کوتاه مدت ۱۵
۶.۲.۲ شبکه‌های عصبی مصنوعی سیام ۱۶
۷.۲.۲ نمایش ویژگی‌ها ۱۷
۸.۲.۲ جاسازی کلمات و مدل سازی احتمالاتی عصبی Word2Vec ۲۰
۹.۲.۲ مسائل پرسش‌ـپاسخ ۲۰
۳.۲ مدل مفهومی ۲۱
۴.۲ جمع بندی ۲۲

۳ روش شناسایی تحقیق ۲۳
۱.۳ مقدمه ۲۳
۲.۳ روش و طرح کلی تحقیق ۲۴
۱.۲.۳ جاسازی پرسش‌ها و پاسخ‌ها ۲۴
۲.۲.۳ شبکه‌های شباهت عمیق ۲۶
۳.۳ پایگاه داده ۲۷
۴.۳ جمع‌بندی ۲۷

۴ نتیجه‌گیری ۲۹
۱.۴ مقدمه ۲۹
۲.۴ پیشنیاز ۳۰
۱.۲.۴ معرفی چند معیار مقایسه ۳۰
۲.۲.۴ نگاهی دقیق‌تر به مسائل دسته‌بندی ۳۰
۳.۲.۴ Accuracy ۳۱
۴.۲.۴ Precision ۳۱
۵.۲.۴ Recall ۳۳
۶.۲.۴ F_1 ۳۳
۳.۴ خلاصه تحقیق ۳۳
۴.۴ بحث در تحقیق ۳۳

۱.۴.۴ جاسازی عبارات ۳۴
۲.۴.۴ شبکه عصبی سیام ۳۹
۵.۴ نتیجه گیری ۴۱
۶.۴ پیشنهادهای برخاسته از تحقیق ۴۱
۷.۴ جمع‌بندی کلی ۴۲

مراجع **۴۵**

فهرست تصاویر

۱.۲ **شبکه‌ی عصبی مصنوعی پیش‌خور** : (۱) لایه‌ی ورودی. (۲) لایه‌های مخفی. (۳) لایه‌ی خروجی ۱۳

۲.۲ **شبکه‌های عصبی مصنوعی بازگشتی**: (۱) مقادیر ورودی (۲) لایه‌های مخفی که تعداد لایه‌ها بیش از یک لایه است. (۳) خروجی (۴) رابطه بازگشتی بین لایه‌های مخفی ۱۴

۳.۲ **شبکه‌ی عصبی مصنوعی LSTM در مقابل شبکه‌ی عصبی مصنوعی بازگشتی** : هر بخش A، به عنوان ورودی مقدار x_t را دریافت و مقدار h_t را به خروجی می‌برد. (آ) ماژول‌های تکرار شونده در شبکه‌های عصبی بازگشتی استاندارد فقط دارای یک لایه هستند. (ب) ماژول‌های تکرار شونده در $LSTM$ها دارای چهار لایه که با هم در تعامل هستند است. ۱۵

۴.۲ **شبکه‌های عصبی مصنوعی سیام**: (۱) لایه‌ی ورودی شبکه که شامل دو دسته ورودی مجزا است. (۲) لایه‌های مخفی شبکه با دو دسته‌ی مستقل ولی با وزن‌های یکسان در شبکه قرار گرفته‌اند. (۳) لایه‌ی شامل استخراج ویژگی برای سنجش میزان شباهت (۴) لایه‌ی بررسی میزان شباهت (۵) لایه‌ی خروجی، شامل امتیاز میزان شباهت دو ورودی. ۱۶

۵.۲ **نمایش بردار ویژگی تنک و بردار ویژگی متراکم** : (۱) جاسازی کلمه. (۲) جاسازی بخشی از گفتار. دو نمونه کدگذاری اطلاعات: کلمه‌ی مورد بحث "Dog" است و کلمه‌ی قبل آن "The" است و "DET" یک $pos-tag$ است. (آ) بردار ویژگی متغیر. هر بعد، نمایشگر یک ویژگی است و ترکیبی از ویژگی‌ها بردار را تشکیل می‌دهد. مقادیر ویژگی باینری هستند. ابعاد بسیار زیاد است. (ب) هر بردار ویژگی، مبتنی بر جاذبه چگال هر ویژگی اصلی ،به عنوان یک بردار نمایش داده می شود. هیچ رمزگذاری صریح ترکیبی از ویژگی‌ها وجود ندارد. ابعاد کم است. نگاشت ویژگی‌ها به یک بردار با اندازه مشخص، بردار هر کلمه را ایجاد می‌کند. ۱۸

۶.۲ **دو معماری مدلسازی روش Word2Vec** : (۱) ورودی شبکه. (۲) طرح ریزی. (۳) خروجی. (آ) معماری $CBOW$ (ب) معماری $Skip-Gram$ . . ۲۰

۱.۳ **معماری Doc2Vec** : (۱) لایه‌ی ورودی (شامل بردار کلمات). (۲) لایه‌ی مخفی(بعلاوه ترکیب و یا میانگین گیری از بردار ایجاد شده برای کلمات، برای ساخت بردار سند). (۳) لایه‌ی خروجی(شامل بردار کلمات). (۴) بردار مربوط به سند . (آ) معماری DM. (ب) معماری $DBOW$. ۲۶

۲.۳ **نمونه‌ای از پرسش_پاسخ درون پایگاه داده** ۲۸

۱.۴ **ماتریس درهم‌ریختگی یا اغتشاش** ۳۱

۲.۴ **نمایش بصری ماتریس درهم‌ریختگی** ۳۲

۳.۴ **مقایسه‌ی چند روش میانگین‌گیری برای ساخت بردار عبارات پرسش و پاسخ از لحاظ معیارهای مختلف :** AM) میانگین حسابی، GM) میانگین هندسی، HM) میانگین هارمونیک، WM) میانگین وزنی. ۳۵

۴.۴ **مقایسه‌ی دو روش ساخت سیستم پرسش پاسخ اتوماتیک از لحاظ معیارهای مختلف:** A) $doc2vec + SNN$، B) SNN + روش جاسازی پیشنهادی. . ۳۸

۵.۴ **مقایسه‌ی دو روش ساخت سیستم پرسش پاسخ اتوماتیک بر اساس نوع شبکه عصبی سیام:** (A شبکه عصبی سیام استاندارد (B شبکه عصبی سیام با وزن‌های مستقل. ۴۰

۶.۴ **نمایی از یک سیستم پاسخگویی** ۴۳

فهرست جداول

۱.۳ مشخصات پایگاه داده ۲۸

۱.۴ بررسی روش‌های مختلف میانگین گیری در ساخت بردار عبارات بر اساس میانگین گیری از بردار کلمات درون عبارات. ۳۴

۲.۴ محاسبه معیارهای مقایسه برای استفاده از میانگین‌گیری‌های متفاوت برای ساخت بردار پرسش‌ها و پاسخ‌ها. ۳۵

۳.۴ نسبت میانگین فاصله‌ی جفت پرسش ـ پاسخ‌ها بر اساس بردار عبارات به روش‌های مختلف میانگین گیری کلمات با روش *doc2vec* ۳۶

۴.۴ بررسی میزان شباهت در میانگین گیری وزنی پویا و میانگین‌گیری وزنی ساده نسبت به روش *doc2vec* ۳۷

۵.۴ بررسی تعداد پاسخ صحیح به سوالات مطرح شده (تعداد کل پرسش‌های مرحله‌ی آزمون ۲۰۰۰ پرسش و مجموع پاسخ‌های مرحله‌ی آزمون ۳۳۰۸ پاسخ می‌باشد). . ۳۷

۶.۴ محاسبه معیارهای مقایسه در دو روش مختلف برای ساخت بردار عبارات . . ۳۸

۷.۴ مقایسه عملکرد دو شبکه‌ی عصبی سیام بصورت شبکه‌ی عصبی سیام استاندارد و شبکه‌ی عصبی سیام تغییر یافته در تعداد تکرار (فاز آموزش شبکه‌ی عصبی) ۶۰۰ دور. ۳۹

۸.۴ محاسبه معیارهای مقایسه در دو روش مختلف استفاده از شبکه‌ی عصبی سیام. . ۴۰

۹.۴ مقایسه روش مقاله‌ی دکتر مینایی و روش پیشنهادی. ۴۱

فصل ۱

مقدمات

۱.۱ مقدمه

از آغاز عصر دیجیتال، پاسخ به سوالات ــ هر گونه سوال ــ بصورت خودکار و با استفاده از رایانه‌ها یکی از رؤیاهای انسان بوده است. درواقع تمایل به برقراری ارتباط با رایانه از طریق گفتگوی انسانی، یکی از انگیزه‌های اصلی ایجاد هوش مصنوعی بوده است. می‌توان از ELIZA [۱] به عنوان شروع واقعی یک تعامل بین انسان و رایانه نام برد.

مسئله‌ی پرسش‌ـ‌پاسخ به عنوان پیدا کردن راه کاری برای پاسخ به پرسش‌های انسان‌ها در موضوعات مختلف بوجود آمد. روش‌های پردازش بر اساس پردازش‌های معنایی و آماری مطرح می‌شوند که هر کدام پیشرفت‌های قابل توجهی داشته‌اند و روز به روز کامل‌تر شده‌اند.

روش‌های پردازش آماری برای پاسخگویی به پرسش‌ها، روش‌های بودند که می‌توانستند با استفاده از ابزار معرفی شده در این شاخه به پیدا کردن قوانین پنهان شده در ساختار زبان‌ها بپردازند که از مهمترین ویژگی‌های این روش‌ها می‌توان به، فارغ از نوع زبان بودن، این روش‌ها اشاره کرد. بنابراین می‌توان با بی توجه بودن به این که با چه زبانی کار می‌کنیم به پردازش آن زبان بپردازیم. برای این کار در گام اول نیاز داشتیم به این که بتوانیم متن را با نمایش مشخصی تبدیل به بردار کنیم و سپس پردازش‌های رایانه‌ای را روی آن‌ها اعمال کنیم.

---

[۱] برنامه‌ی رایانه‌ای که اقای Joseph Weizenbaum آن را در سال ۱۹۶۶ طراحي کرد تا هوشمند بودن رایانه را نشان دهد.

در این زیر شاخه از علم نیز روش‌های زیادی بوجود آمدند که روش‌های خودکار با استفاده از شبکه‌های عصبی مصنوعی، از جمله‌ی روش‌های جدید این شاخه از علم بودند. در طی سال‌های اخیر روش‌هایی تحت عنوان مدل سازی احتمالاتی عصبی توسط شرکت‌های مختلف ساخته شده‌اند که کلمات را به بردارهای معناداری تبدیل می‌کردند و این امر منجر به ساخت و استفاده از روش‌های قوی‌تری در پردازش زبان‌های طبیعی شد.

ساختار فصل

در بخش ۱-۱ به بررسی مختصری از پردازش زبان‌های طبیعی و شاخه‌ی پرسش‌ـ‌پاسخ آن، می پردازیم، سپس در بخش ۱-۲ برخی از مسائلی که روند پیشبرد آنها به واسطه‌ی محدودیت‌های ایجاد شده توسط حل نشدن مساله‌ی پرسش‌ـ‌پاسخ، مختل شده است و برخی از کارهایی که منجر به تلاش برای حل چنین مسائلی می شود، می‌پردازیم، در بخش ۱-۳ هدف پایان‌نامه را در جهت حل مسأله مشخص شده مطرح می‌کنیم و اهداف اصلی پایان‌نامه را متذکر می‌شویم. در انتها مرور کلی از ساختار پایان‌نامه در بخش ۱-۴ ارائه می‌گردد.

۲.۱ یک زمینه ی تحقیقات طولانی

تحقیق در پاسخ به سوال در اوایل دهه‌ی ۱۹۶۰ آغاز شد و از آن به بعد، این عملیات تحت عنوان مسئله‌ی پرسش‌ـ‌پاسخ اهداف گسترده‌تری را در زمینه‌های بازیابی اطلاعات، زبان‌شناسی کاربردی، تعامل با رایانه و ارتباطات انسانی، بوجود آورد. این بخش خلاصه‌ای مختصر در مورد تاریخ مسئله‌ی پرسش‌ـ‌پاسخ است و انگیزه‌ی کار در این پایان‌نامه را ارائه می‌دهد.

۱.۲.۱ آغاز مسأله‌ی پرسش‌ـ‌پاسخ

کارهای اولیه در زمینه پاسخ به پرسش‌های مربوط به دامنه‌های بسیار محدود بود و شامل بازیابی اطلاعات از پایگاه داده‌های کوچک، مانند سوابق رویدادهای ورزشی [۲] می‌شد.

یکی از اولین الگوریتم‌های پاسخ دادن به پرسش‌های عمومی، که در سال‌های ۱۹۷۰ توسط سیمونز ارائه شد [۳] ، شامل مراحل زیر است: ابتدا مجموعه ای از اسناد را برای حل مسئله‌ی پرسش‌ـ‌پاسخ، یک پایگاه داده از ساختارهای معنایی که معانی جمله‌ها را تشکیل می‌دهند را جمع‌آوری کردند. سپس مجموعه‌ای از ساختارهایی که چندین مفاهیم واژگانی را با این سوال به اشتراک می‌گذارد انتخاب شدند تا لیستی از پاسخ‌های کاندید را ایجاد کنند. در نهایت، این سوال در برابر هر ساختار از نامزدها قرار داشت و بهترین نامزد به عنوان پاسخ انتخاب می‌شد.

این رویکرد بسیار ساده، نقش مهم پردازش معنایی را نشان می‌دهد، از اطلاعات غیر از موارد موجود در سیستم‌های پایگاه داده بهره می‌گیرد و از بازیابی اطلاعات متمایز است. این روال تا ۱۹۸۰ مرسوم بود و شرکت‌های زیادی از آن استفاده کردند که از شرکت‌های معروف که در محصولات خود استفاده کردند می‌توان به شرکت یونیکس و استفاده‌ی از این روال در

مشاور یونیکس این شرکت، اشاره کرد. [۴]

اگر چه به نظر می رسد که آن‌ها (سیستم‌هایی با ساختار مشاور یونیکس) همیشه به درد نخور بوده‌اند ولی، این سیستم‌ها به توسعه نظریه‌های زبان‌شناسی محاسباتی [۲] و استدلال کمک کرده‌اند.

پرسش_پاسخ دامنه باز در اواخر ۱۹۹۰ مطرح شد و طولی نکشید که جزء ای از استانداردهای مسئله‌ی پرسش_پاسخ شد. در پرسش_پاسخ دامنه باز، محدوده سوالات احتمالی محدود نمی‌شود، بنابراین چالش بسیار سنگین‌تر بر روی سیستم قرار می‌گیرد، زیرا کامپایل کردن همه ساختارهای معنایی ممکن که در یک متن وجود دارد، غیرممکن است.

توانایی درک و ارتباط در زبان‌های طبیعی؛ مانند انگلیسی یا فارسی، چالش دیگری بود که این شاخه از علم درگیر آن شد. پیدا کردن راه کاری برای ارتباط با زبان‌های مختلف از جمله زمینه‌هایی بود که این شاخه از علم برای آن بوجود آمد ولی با توجه به ساختارهای معنایی مختلفی که برای تمام زبان‌های طبیعی وجود داشت این کار را ناشدنی‌تر می کرد.

در انتهای سال‌های ۱۹۹۰ شروع روش‌های آماری برای حل مسائل پردازش زبان بود و با استفاده از این روش‌ها، که بر پایه‌ی شمارش کلمات درون متون ساخته شده بودند، می‌توانستیم با استفاده از محاسبات ریاضی و آماری به پردازش زبان‌های طبیعی و البته مسائل پرسش_پاسخ بپردازیم .

با ظهور یادگیری عمیق و ورود روش‌های پردازش شبکه‌های عصبی مصنوعی عمیق در پردازش متن و البته مسائل پرسش_پاسخ، پیدا کردن ساختارهای معنایی به شبکه‌های عصبی مصنوعی عمیق سپرده شد و در این روش‌ها می‌توانیم با آموزش شبکه‌های عصبی مصنوعی به کشف روابط مخفی درون متون بپردازیم و سیستم‌های قدرتمندی را برای پاسخگویی به سوالات آموزش دهیم این تنها قسمتی از جذابیت‌های این روش است و از دیگر جذابیت‌های این روش می‌توان به فارغ از زبان بودن این روش‌ها اشاره کرد و این به آن معناست که می‌توان از الگوریتم‌های یکسان می‌توان در پردازش زبان‌های طبیعی گوناگون استفاده کرد.

۲.۲.۱ برخی از مسائل قابل بحث در زمینه ی پرسش_پاسخ

همان‌طور که قبلا اشاره شد، به طور گسترده‌ای پذیرفته شده است، که پاسخ به پرسش یک گام فراتر از بازیابی اطلاعات استاندارد است. علی‌رغم داشتن یک نظریه‌ی طولانی تحقیق، پرسش_پاسخ هنوز مجموعه‌ای از مسائل حل نشده را در بتن خود دارد که در زیر در این راستا بحث می شود.

---

[۲] *computational linguistics*

۳.۱ بیان مسأله

مسئله‌ی پرسش‌ـ‌پاسخ به عنوان پیدا کردن راه‌کاری برای پاسخ به پرسش‌های انسان‌ها در موضوعات مختلف مطرح می‌شود. مشکلاتی از قبیل تنوع زبانی، مشکلات دستوری در مکالمات غیر رسمی انسانی، تنوع در کلمات و جملات، حضور احساسات در لابه‌لای عبارات و ... ، پردازش زبان‌های طبیعی را به سوی الگوریتم‌های کاراتر برای پاسخگویی خودکار به پرسش‌ها برده است. در این روش‌ها از عملیاتی، فارغ از روش‌های مبتنی بر ساختارهای دستوری، استفاده می‌شود.

در این پایان نامه با استفاده از یادگیری عمیق سعی در پیدا کردن پاسخ سوالات با استفاده از پیدا کردن شباهت‌های معنایی بین جفت پرسش و پاسخ می‌کنیم.

۱.۳.۱ سیستم‌های پرسش‌ـ‌پاسخ خودکار

برنامه‌های رایانه‌ای که ساخته می‌شوند تا بتوانند با استفاده از دانش و یا پایگاه داده‌ای که در اختیارشان قرار می‌گیرد به پرسش‌هایی که از جانب کاربران این برنامه‌های رایانه‌ای، برایشان مطرح می‌شود، پاسخ دهند، تحت عنوان سیستم‌های پرسش‌ـ‌پاسخ خودکار، قرار می‌گیرند.

این سیستم‌ها با بازیابی اطلاعات از دانشی که در اختیار دارند الگوهایی را برای پیدا کردن پاسخ‌ها بدست می‌آورند. گاهی بصورت نظارت شده به آنها آموزش داده می‌شود و گاهی بصورت روش‌های غیرنظارتی سیستمی ساخته می‌شود تا خود به کشف ارتباطات بین پرسش‌ها و پاسخ‌ها بپردازد. از جمله‌ی سیستم‌های پرسش‌ـ‌پاسخ مشهور شرکت‌های بزرگ می‌توان به سیستم‌هایی چون موتور جستجوی گوگل، سیری [۳] شرکت اپل و یا کرتانا [۴] محصول مایکروسافت اشاره کرد.

۲.۳.۱ پیدا کردن شباهت بین جفت پرسش‌ـ‌پاسخ

برای پاسخگویی به پرسش مطرح شده یک روش آن است که به دنبال پیدا کردن متن یا عباراتی در پایگاه داده بگردیم که شبیه ترین متن به پرسشمان باشد تا شاید بتواند پاسخ مورد نظر باشد یا بتواند به پاسخ گفتن به پرسش کمک کند. این شباهت‌ها می‌تواند تعداد کلمات مشابه درون متون با پرسش باشد و یا کمترین فاصله دو بردار مربوط به متون و پرسش باشد.

---

[۳] $Siri$

[۴] $Cortana$

۳.۳.۱ استفاده از یادگیری عمیق

خوشبختانه دانشمندان فعال در این شاخه از علم توانسته‌اند با استفاده از روش‌های مبتنی بر یادگیری عمیق کلمات و عبارات را به شیوه‌های متنوع،ی بسته به نوع کاربرد تبدیل به بردارهای عددی کنند. این دسته از فعالیت‌ها منجر به ساخت بردارهای عددی معناداری در فضاهای عددی شده است که پردازش متون را گاهی بسیار ساده‌تر کرده است.

۴.۱ اهمیت و ضرورت تحقیق

پردازش زبان‌های طبیعی و منحصرا پردازش مسئله‌ی پرسش_پاسخ می‌تواند در زمینه‌های فراوانی به انسان در راحت‌تر زندگی کردن کمک کند لذا سعی در پیشبرد آن، علاوه بر شیرین بودن خود موضوع، می‌تواند یک عمل بشر دوستانه به شمار بیاید.

پیدا کردن روش‌های قوی‌تر در پیدا کردن شباهت بین پرسش‌ها و پاسخ‌ها می‌تواند سیستم پرسش_پاسخ را سریع‌تر و دقیق‌تر آموزش دهد.

۵.۱ اهداف

هدف از این پایان نامه معرفی روش برای پیدا کردن هرچه سریع‌تر پاسخ در سیستم‌های پرسش_پاسخ است و در این پایان نامه روشی ارائه می‌شود که می‌تواند سیستم پرسش_پاسخ خودکار مطرح شده در این پایان‌نامه را سریع‌تر و دقیق‌تر آموزش دهد.

۶.۱ تعریف‌های اولیه

۱.۶.۱ پردازش زبان‌های طبیعی

پردازش زبان طبیعی به معنای مطالعه بر روی درک رایانه‌ها از گفتار، نوشتار، و ارتباط انسانی می باشد. رایانه‌ها با مجهز شدن به این درک، می‌توانند تحلیل احساسی یا دیگر تحلیل‌های متنی را در مقیاس بزرگ انجام دهند و داده‌های معنادار فراهم آورند.

۲.۶.۱ سیستم‌های بازیابی اطلاعات

سیستم های بازیابی اطلاعات (IR) سیستم‌هایی هستند که وظیفه بازیابی اسناد مرتبط با پرسش کاربر را بر عهده دارند. این سیستم‌ها پرسش کاربرد را در فرم خاصی دریافت می‌کنند و از میان منابع موجود اسناد مرتبط را استخراج می‌کنند.

۳.۶.۱ مسأله‌ی پرسش_پاسخ

در این مسائل، سوال به زبان طبیعی و بدون هیچ محدودیت معنایی و نحوی به عنوان ورودی به سیستم داده می‌شود. وظیفه سیستم یافتن جوابی دقیق، کوتاه و کامل برای سوال داده شده در کوتاه‌ترین زمان ممکن است. به این منظور سیستم‌های QA تکنیک‌های بازیابی اطلاعات، استخراج اطلاعات و پردازش زبان طبیعی را با هم به کار می‌گیرند.

۴.۶.۱ شبکه‌های عصبی مصنوعی

یک شبکه عصبی مصنوعی ایده‌ای برای پردازش اطلاعات است که از سیستم عصبی زیستی الهام گرفته و مانند مغز به پردازش اطلاعات می‌پردازد. عنصر کلیدی این ایده، ساختار جدید سیستم پردازش اطلاعات است. این سیستم از شمار زیادی عناصر پردازشی بهم پیوسته به نام نورون‌ها تشکیل شده که برای حل یک مسئله با هم هماهنگ عمل می‌کنند.

شبکه‌های عصبی مصنوعی با پردازش داده‌های تجربی، دانش یا قانون نهفته در ورای داده‌ها را به ساختار شبکه منتقل می‌کند که به این عمل یادگیری می‌گویند. اصولاً توانایی یادگیری مهم‌ترین ویژگی یک سیستم هوشمند است. سیستمی که قابلیت یادگیری داشته باشد، منعطف‌تر است و ساده‌تر برنامه‌ریزی می‌شود، بنابراین بهتر می‌تواند در مورد مسائل و معادلات جدید پاسخگو باشد.

۵.۶.۱ یادگیری عمیق

یادگیری عمیق شاخه ای از بحث یادگیری ماشین و هوش مصنوعی و مجموعه ای از الگوریتم هایی است، که تلاش می کنند، مفاهیم انتزاعی سطح بالا را با استفاده از یادگیری در سطوح و لایه های مختلف مدل کنند.

۶.۶.۱ مدل‌سازی احتمالاتی عصبی

استفاده از شبکه‌های عصبی با یک لایه‌ی مخفی که به مدل‌سازی کلمات، عبارات، جملات و اسناد می‌کند و روش‌های گوناگونی وجود دارد که می‌تواند این مدل‌سازی را بر اساس قوانین معنایی [۵] و یا دستوری [۶] مدل کند.

---

[۵] روش‌های ساخته شده توسط شرکت گوگل با عناوینی چون : $Word2vec$ ، $Doc2vec$ و ... .

[۶] روش‌های ساخته شده توسط گروه $Stanford\ NLP$ با عنوان : $GloVe$.

۷.۶.۱ شبکه‌های عصبی مصنوعی شباهت عمیق

شبکه‌های عصبی چندلایه‌ی عمیق که سعی در پیدا کردن ویژگی‌های مشابه دو شی را دارند و عملیات انتخاب ویژگی را درون لایه‌های پنهان انجام می‌دهند و بعد از پیدا کردن و انتخاب ویژگی‌ها به مقایسه‌ی آن می‌پردازند. در درون این شبکه‌ها ممکن است از فیلترهایی برای انتخاب ویژگی‌ها استفاده شود و یا این که شبکه با شیوه‌های غیر نظارتی یاد بگیرد که چه ویژگی‌هایی را بایستی انتخاب کند.

۷.۱ جمع بندی

در انتهای این فصل آموختیم که پردازش زبان‌های طبیعی می‌تواند به عنوان شاخه‌ی ارزشمندی در علوم کامپیوتر و چه بسا تمام علوم باشد و همچنین سیستم‌های پرسش_پاسخ نیز که یکی از دلایل وجود چنین علمی بوده‌اند، نگاه ویژه‌ی محققین به خود را در علوم جدید دارند و هنوز از جمله‌ی مسائلی در دنیا محسوب می‌شود که علاوه بر محبوبیت بالا در میان دانشمندان علوم کامپیوتری، جای کار فراوانی را دارند. در ادامه دریافتیم که ابزارهای یادگیری عمیق نیز توانسته‌اند جایگاه خود را در ارائه‌ی روش‌های حل کارا، داشته باشند.

در انتها قرار بر این شد تا ما نیز از این یادگیری عمیق در تقویت سیستم‌های پاسخگویی موجود، استفاده کنیم و عملکرد این سیستم‌ها را تا حدودی بهبود بدهیم. در ادامه به تفضیل به ارائه‌ی روشمان می‌پردازیم.

فصل ۲

ادبیات و پیشینه تحقیق

۱.۲ مقدمه

برای مدت زمان طولانی، تکنیک‌های هسته‌ی پردازش زبان‌های طبیعی تحت تأثیر روش‌های یادگیری ماشین بود که از مدل‌های خطی مانند ماشین‌های بردار پشتیبانی [۱] یا رگرسیون لجستیک [۲] استفاده می‌کردند، آموزش بر روی بردارهای ویژگی که بسیار بزرگ و تنک [۳] بودند انجام می شد.

به تازگی، در این زمینه موفق به تغییر در روند ساخت بردارهای ویژگی، از این چنین مدل‌های خطی با ورودی‌های تنک، به مدل‌های غیر خطی شبکه عصبی با ورودی‌های متراکم [۴] شده‌اند و به این صورت استفاده از بسیاری از تکنیک‌های شبکه عصبی را آسان کرده‌اند، گاهی اوقات برای جایگزین کردن تکنیک‌های شبکه عصبی با الگوریتم‌ها و روش‌های قدیمی خطی، یکی از مشکلات اصلی، ورودی‌هایی غیر قابل استفاده، بودند. در ادامه به بررسی روند شکل‌گیری اصول استفاده از تکنیک‌های شبکه عصبی در مسائل پردازش زبان‌های طبیعی و منحصراً مسئله‌ی پرسش‌ـ‌پاسخ می‌پردازیم.

---

[۱] *support vector machines (SVM)*

[۲] *logistic regression*

[۳] *sparse*

[۴] *dense*

۲.۲ مروری بر تحقیقات پیشین

هر چند بسیاری از مطالب مورد بحث در رابطه با استفاده از شبکه‌های عصبی در مسائل پردازش زبان‌های طبیعی است ولی سعی شده است تمرکز بیشتر بر روی مسئله‌ی پرسش‌ـ پاسخ باشد و موضوعات به‌طور خاص از این دیدگاه مورد بررسی قرار گیرد.

۱.۲.۲ پردازش زبان‌های طبیعی

فرآیند پردازش زبان‌های طبیعی [۵]، به عنوان زبان شناسی محاسباتی [۶] هم شناخته می‌شود. این شاخه از علم، محدوده‌ی تحقیقاتی است که توسط تعدادی از حوزه‌های مختلف ازجمله: زبان شناسی، روانشناسی، فلسفه، علوم‌شناختی، آمارواحتمال و یادگیری‌ماشین شناخته شده است. این علم شامل مهندسی مدل‌های کامپیوتری و فرآیندهایی برای حل مشکلات عملی در درک زبان‌های انسانی است. درنهایت این راه حل‌ها برای ساخت نرم افزارها مفید است و می‌تواند در کاربردهای مختلف استفاده شود.

کار در NLP از دیدگاهی، می‌تواند به دو بخش کلی تقسیم شود: شاخه‌های اصلی و برنامه‌های کاربردی. شاخه‌های اصلی، مسائل اساسی، مانند مدل سازی زبان را مطرح می‌کنند که نشان می‌دهد که ارتباطات کیفیتی [۷] بین کلمات طبیعی وجود دارد. پردازش هستی شناسی [۸] ، و یا تقابل با تقسیم مولفه های معنی‌دار کلمات و شناسایی بخش‌های واقعی سخنرانی کلمات [۹] به عنوان پردازش‌های شاخه‌ی اصلی قرار می‌گیرند. همچنین پردازش نحوی [۱۰] ، که قوانین حاکم را به عنوان پردازش‌هایی که منجر به پردازش‌های معنایی می‌شود و تلاش می‌کند تا معنای کلمات، عبارات، و اجزای سطح بالاتری را در متن بدست آورد. زمینه‌های کاربردی نیز شامل موضوعاتی نظیر استخراج اطلاعات مفید از اسناد و مدارک، ترجمه متون، خلاصه‌ای از آثار نوشته شده، پاسخ خودکار پاسخ به پرسش‌ها [۱۱] با درنظر گرفتن پاسخ‌های احتمالی، طبقه‌بندی و خوشه‌بندی اسناد، می‌باشد.

تلاش‌های اولیه در NLP اغلب بر پایه قوانین دستوری زبان‌ها بوده است، گاهی اوقات این قوانین به صورت مجزا در نظر گرفته می‌شد و از آن‌ها برای حل مشکلات خاص در جای مناسب، به کار گرفته می‌شد. [۵۲، ۵۳]

پردازش مبتنی بر اساس قوانین دستوری به نحوی کارا نیست، زیرا زمانی که انسان‌ها می‌نویسند و یا صحبت می‌کنند، اغلب از تمام دستوراتی که برای استفاده‌ی صحیح از زبان

---

[۵] $natural\ language\ processing (NLP)$
[۶] $computational\ linguistics$
[۷] $quantifying$
[۸] $morphological$
[۹] $speech\ of\ words$
[۱۰] $syntactic$
[۱۱] $question - answering\ systems$

تعریف می‌شود، اطاعت نمی کنند. علاوه بر این، با گسترش رسانه‌های اجتماعی، زبان‌های غیر رسمی استفاده می‌شوند، که عمدا دستورات به درستی در مورد املا و دستور زبان رعایت نمی‌شوند. درنتیجه برای برآوردن نیاز مدیریت زبان بیش از آنچه که به نظر می‌رسد، نیاز است. در دهه‌ی ۱۹۹۰ استفاده از روش‌های آمار و احتمالاتی همراه با الگوریتم‌های یادگیری ماشین برای پردازش‌های زبانی به کار گرفته شد. [۵، ۵۴، ۵۵]

با گذشت زمان، تعدادی از روش های یادگیری ماشین مانند: $nave\ Bayes$، Kـ نزدیکترین همسایه، [۱۲] مدل‌های مخفی مارکوف، [۱۳] زمینه‌های تصادفی شرطی، [۱۴] درخت‌های تصمیم، [۱۵] جنگل‌های تصادفی [۱۶] و ماشین های بردار پشتیبانی در NLP به طور گسترده ای مورد استفاده قرار گرفت. با این حال، در طی چند سال گذشته، در NLP تغییر عمده‌ای صورت گرفته است و بسیاری از این روش‌ها با استفاده از مدل‌های عصبی به طور کامل جایگزین شده‌اند.

۲.۲.۲ شبکه‌های عصبی مصنوعی و پردازش زبان‌های طبیعی

شبکه های عصبی مصنوعی مدل‌های یادگیری قدرتمندی هستند. ما در این قسمت در مورد دو دسته معماری شبکه عصبی مصنوعی بحث خواهیم کرد، که درواقع بر پایه‌ی شبکه‌های عصبی مصنوعی پیش خور [۱۷] و شبکه‌های عصبی مصنوعی بازگشتی و مجازی [۱۸] می‌باشند. شبکه‌های عصبی مصنوعی پیش خور شامل شبکه‌هایی با لایه‌های کاملا متصل [۱۹] همچون پرسپترون چند لایه [۲۰] و شبکه‌هایی با لایه‌های کانولوشن [۲۱] و لایه‌های جمع‌آوری [۲۲] باشند. همه‌ی شبکه‌ها عملیات طبقه‌بندی [۲۳] را هر کدام با نقاط قوت مختلف انجام می‌دهند.

شبکه‌های عصبی مصنوعی پیش‌خور به طور کامل متصل یادگیرنده‌های غیر خطی هستند که عمدتا می‌توانند به عنوان جایگزینی در هر جا که یک یادگیرنده‌ی خطی استفاده می‌شود، استفاده شود. این شامل مسائل طبقه‌بندی دوتایی و چند طبقه و مسائل پیش‌بینی با ساختاری پیچیده‌تر، است.

غیر خطی بودن شبکه، و همچنین توانایی ادغام کلمه‌های تعبیه شده، قبل از آموزش،

---

[۱۲] $k-nearest\ neighbors$
[۱۳] $hidden\ markov\ models$
[۱۴] $conditional\ random\ fields$
[۱۵] $decision\ trees$
[۱۶] $random\ forests$
[۱۷] $feed\ forward\ neural\ network$
[۱۸] $recurrent\ /\ recursive\ neural\ networks$
[۱۹] $fully\ connected\ layers$
[۲۰] $multi-layer\ perceptron$
[۲۱] $convolutional\ layer$
[۲۲] $pooling\ layers$
[۲۳] $classifiers$

اغلب به دقت طبقه بندی مرتبط است. تعدادی از دانشمندان این حوزه، موفق به استفاده از روش‌هایی شدند که با جایگزینی شبکه به طور کامل متصل پیش‌خور با یک مدل یادگیرنده‌ی خطی، موفق به کسب نتایج بهبود یافته‌ای شدند. [۶، ۷، ۸، ۹] همچنین در شاخه‌های مختلف پردازش زبان‌های طبیعی شبکه‌های عصبی مصنوعی پیشخور جایگزین مناسبی برای طبقه بندها بودند [۱۰، ۱۱، ۱۲] که معمولا همراه با استفاده از بردارهای کلمات پیش آموزش یافته، بودند. [۱۳، ۱۴] و همچنین مطالعاتی که در حوزه‌ی طبقه‌بندی پرسش‌ها و پاسخ‌ها در مسائل پرسش_پاسخ انجام شده‌اند. [۱۵]

شبکه هایی با لایه‌های کانولوشن و لایه‌های جمع آوری [۲۴] برای مسائل طبقه‌بندی، که در آن انتظار می‌رود سرنخ‌های محلی قدرتمندی در مورد عضویت در کلاس پیدا کنیم، در حالی که ممکن است این سرنخ‌ها در مکان‌های مختلف در ورودی ظاهر شود. به عنوان مثال، در یک مسئله‌ی طبقه بندی سند، یک عبارت کلیدی می‌تواند در تعیین موضوع سند [۱۶] کمک کند. ما می‌خواهیم به یاد بیاوریم که برخی از ترفندهای کلمات شاخص خوبی از موضوع هستند و لزوما گاهی در سند ظاهر نمی‌شود. لایه‌های کانولوشن و لایه‌های جمع‌آوری این اجازه را می‌دهد تا مدل یاد بگیرد چنین شاخص‌های محلی را بدون در نظر گرفتن موقعیت خود می‌تواند پیدا کند.

در زبان طبیعی ما اغلب با داده‌های ساختاری از اندازه غیر یکسان، مانند توالی‌ها [۲۵] کار می‌کنیم. ما می‌خواهیم که بتوانیم نظم و انضباط را در چنین سازه‌هایی به دست آوریم یا شباهت‌های بین چنین ساختاری را مدل کنیم. در بسیاری از موارد، این به معنی رمزگذاری ساختار به عنوان یک بردار با طول ثابت است، که پس از آن می‌تواند برای یادگیری آماری دیگری، مورد پردازش قرار گیرد. با وجود آن که معماری‌های کانولوشن به ما این اجازه را می‌دهد که مقادیر دلخواه زیادی را به عنوان بردارهای با طول ثابت در نظر بگیریم، که ویژگی‌های [۲۶] برجسته‌ای‌شان ر،ا با از بین بردن اطلاعات ساختاری زیادی به نمایش می‌گذارند. از سوی دیگر، معماری‌های مجازی و بازگشتی [۲۷] ، ما را قادر می‌سازد تا با توالی‌ها و درخت‌ها کار کنیم در حالی که بسیاری از اطلاعات ساختاری را حفظ می‌کنیم. شبکه‌های مجازی [۱۷] برای مدل سازی توالی‌ها طراحی شده‌اند، و همچنین شبکه‌های بازگشتی [۱۸] که تعمیم شبکه‌های مجازی هستند، می‌توانند درختان را مدیریت و پردازش کنند.

مدل‌های شبکه‌های عصبی مصنوعی مجازی نشان داده‌اند که نتایج بسیار خوبی برای مدل سازی زبان [۱۹، ۲۰، ۲۱، ۲۲، ۲۳، ۲۴، ۲۵] و همچنین برای برچسب گذاری توالی‌ها [۲۶، ۲۷] ارائه می‌دهند. مدل‌های بازگشتی نیز در پژوهش‌های مربوط به پرسش_پاسخ نیز بسیار موثر بودند.

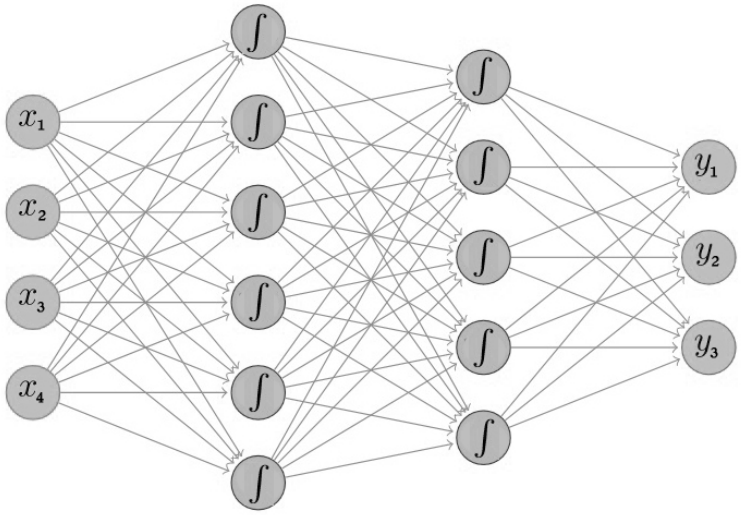
---

[۲۴] *pooling layer*

[۲۵] *sequences*

[۲۶] *features*

[۲۷] *recurrent and recursive architectures*

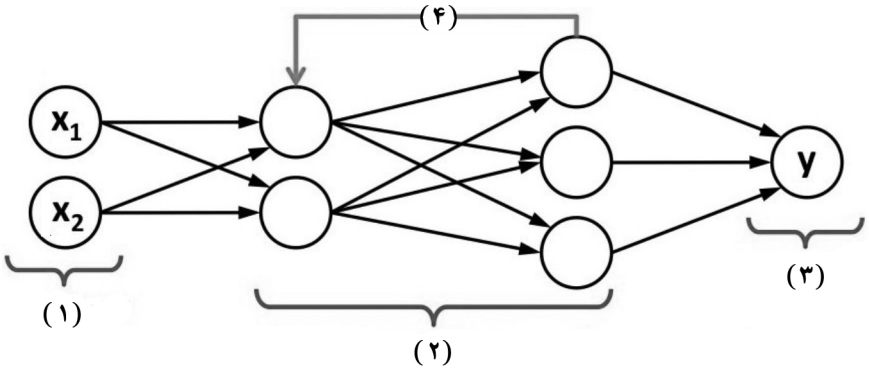


شکل ۱.۲: **شبکه‌ی عصبی مصنوعی پیش‌خور** : (۱) لایه‌ی ورودی. (۲) لایه‌های مخفی. (۳) لایه‌ی خروجی

۳.۲.۲ شبکه‌های عصبی مصنوعی پیش‌خور

شبکه‌های عصبی، همان‌طور که از نام آن پیداست الهام گرفته از مغز انسان می‌باشند. شبکه‌های عصبی بر اساس مکانیسم محاسبات مغز انسان، که از واحدهای محاسباتی به نام نورون تشکیل شده است، کار می‌کنند. یک نورون یک واحد محاسباتی است که ورودی‌ها و خروجی‌های عددی دارد. هر ورودی دارای وزن مرتبط است. نورون ورودی، به وسیله ضرب در وزن متناظر با آن و اعمال یک تابع غیر خطی بر روی آن خروجی را می‌سازد. نورون‌ها به یکدیگر متصل می‌شوند و تشکیل یک شبکه عصبی مصنوعی را می‌دهند. خروجی یک نورون می‌تواند به ورودی‌های یک یا چند نورون دیگر، به عنوان ورودی اعمال شود. چنین شبکه‌هایی نشان داده‌اند که دستگاه‌های محاسباتی بسیار توانایی، را تشکیل می‌دهند. اگر وزن‌ها به درستی تنظیم شوند، یک شبکه عصبی با نورون‌های کافی و یک تابع فعال غیرخطی می‌تواند طیف گسترده‌ای از توابع ریاضی را تقریب بزند.

یک شبکه عصبی پیش‌خور معمولی ممکن است، به صورت شکل ۱.۲ ساخته شده باشد. هر دایره یک نورون است و پیکان‌های ورودی آن، ورودی‌های نورون است و پیکان‌های خروجی، خروجی‌های نورون هستند. هر پیکان دارای وزنی است، که اهمیت آن را نشان می‌دهد (که به دلیل شلوغ شدن شکل، در شکل نشان داده نشده است). نورون‌ها در درون لایه‌ها قرار می‌گیرند و جریان اطلاعات را بازتاب می‌دهند. لایه‌ی ابتدایی دارای پیکان‌های ورودی نیست و نمایشگر ورودی‌های شبکه است. لایه‌ی انتهایی دارای پیکان‌های خروجی



شکل ۲.۲: **شبکه‌های عصبی مصنوعی بازگشتی**: (۱) مقادیر ورودی (۲) لایه‌های مخفی که تعداد لایه‌ها بیش از یک لایه است. (۳) خروجی (۴) رابطه بازگشتی بین لایه‌های مخفی

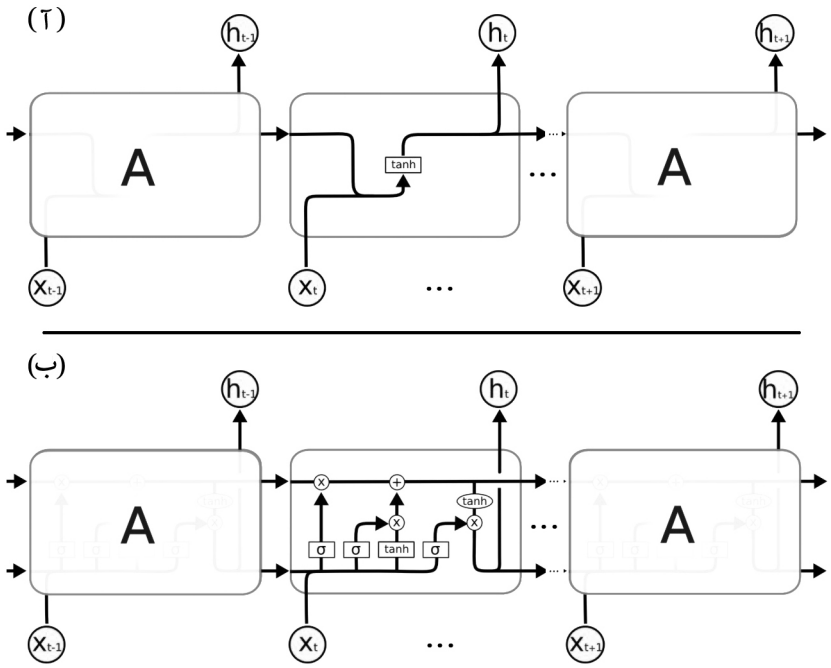
نیست و خروجی شبکه است. لایه‌های دیگر ”لایه‌های پنهان [۲۸]“ در نظر گرفته می‌شوند. شکل سیگموئید داخل نورون‌ها در لایه‌های میانی یک تابع غیرخطی (که معمولا $۱/(۱+e^x)$ است و قبل از انتقال مقدار نورون به خروجی بر روی آن اعمال می‌شود. در این شبکه، هر نورون به تمام نورون‌ها در لایه بعدی متصل می‌شود.

۴.۲.۲ شبکه‌های عصبی مصنوعی بازگشتی

شبکه های عصبی مصنوعی بازگشتی [۲۹] بطور خاص برای پردازش داده‌های سری یا دنباله‌دار مفید هستند و در آن‌ها هر نورون یا واحد پردازشی قادر به حفظ حالت داخلی یا یا همان حافظه به منظور حفظ اطلاعات مرتبط با ورودی قبلی می‌باشد. این ویژگی بطور ویژه در کاربردهای مختلف مرتبط با داده‌های سری اهمیت اساسی پیدا می‌کند. به واسطه وجود ارتباطات دنباله‌ای در زبان‌های طبیعی، این شبکه‌ها در بسیاری از مسائل مربوط به پردازش زبان‌های طبیعی به کار گرفته می‌شوند. حفظ حالت درونی یا همان قابلیت حافظه، به شبکه کمک می‌کند تا قادر به فهم و کشف ارتباط بین لغات مختلف در دنباله‌های طولانی تر باشد. لازم به ذکر است که ما انسان‌ها نیز زمانی که یک جمله را می‌خوانیم با توجه به زمینه محتوایی که هر کلمه در آن قرار گرفته است معنای آن را استنباط می‌کنیم. به عبارت بهتر با توجه به کلمات قبلی و (در بعضی موارد حتی بعدی) زمینه محتوایی را استنباط کرده و با توجه به آن معنای یک کلمه را درک می‌کنیم .

ایده اصلی پشت این نوع از معماری بهره‌برداری از این ساختار سری داده است. نام این شبکه عصبی از این واقعیت بدست می‌آید که این نوع از شبکه‌ها بصورت بازگشتی عمل می‌کنند شکل ۲.۲. یعنی یک عملیات برای تک تک المان‌های یک دنباله (کلمه،جمله،...) انجام

[۲۸] *hidden layer*
[۲۹] *recurrent artificial neural networks*



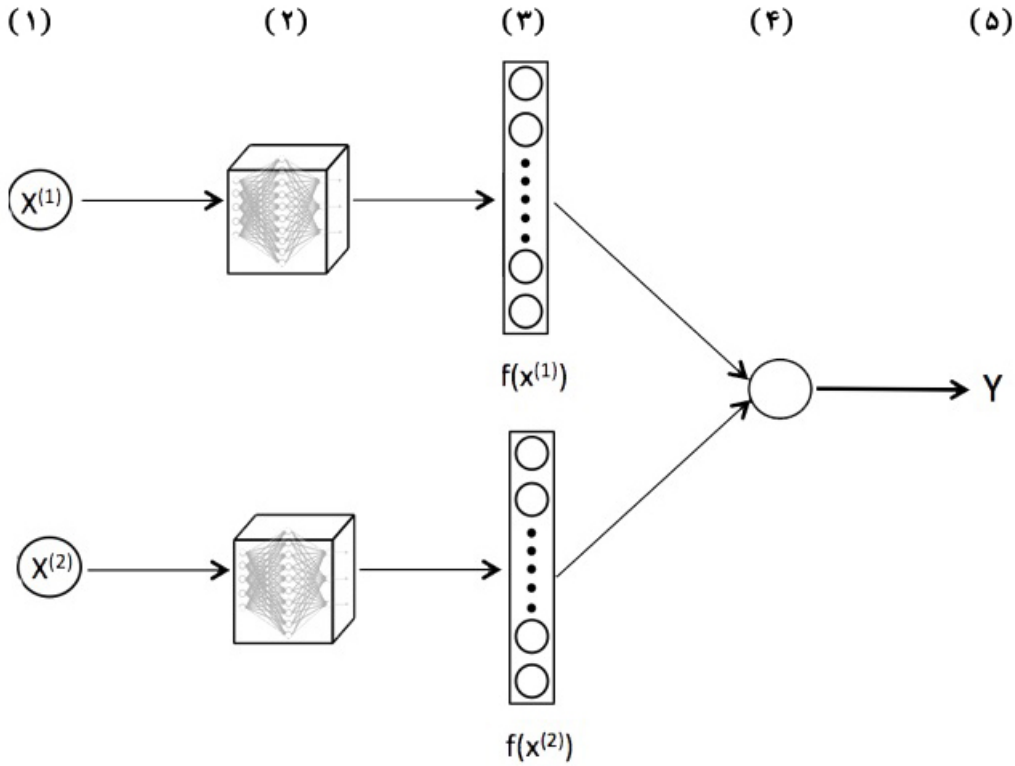
شکل ۳.۲: **شبکه‌ی عصبی مصنوعی LSTM در مقابل شبکه‌ی عصبی مصنوعی بازگشتی** : هر بخش A، به عنوان ورودی مقدار x_t را دریافت و مقدار h_t را به خروجی می‌برد. (آ) ماژول‌های تکرار شونده در شبکه‌های عصبی بازگشتی استاندارد فقط دارای یک لایه هستند. (ب) ماژول‌های تکرار شونده در $LSTM$ها دارای چهار لایه که با هم در تعامل هستند است.

می‌گیرید و خروجی آن وابسته به ورودی فعلی و عملیات‌های قبلی است. این مهم از طریق تکرار یک خروجی از شبکه در زمان t با ورودی شبکه در زمان $t+1$ انجام می‌شود. (یعنی خروجی از مرحله قبل با ورودی تازه در مرحله جدید ترکیب می‌شود.) این چرخه ها اجازه وجود اطلاعات از یک گام زمانی به گام زمانی بعدی را موجب می‌شوند. به عبارت بهتر این نوع شبکه‌ها دارای حلقه‌ای در درون خود هستند که به وسیله‌ی آن می‌توانند اطلاعات را در حین خواندن ورودی از نورون‌ها عبور دهند.

۵.۲.۲ شبکه‌های عصبی مصنوعی حافظه طولانی کوتاه مدت

شبکه‌های عصبی مصنوعی حافظه طولانی کوتاه مدت [۳۰] یک نوع مدل یا ساختار برای داده‌های ترتیبی است که برای توسعه شبکه‌های عصبی بازگشتی RNN ظهور کرد. عبارت "$long\ term\ memory$" به وزن‌های یادگرفته شده و "$short\ term\ memory$" به حالت‌های درونی سلول‌ها اطلاق می‌شود. $LSTM$ برای حل مشکل پدیده ناپدید شدن گرادیان در شبکه‌های عصبی بازگشتی، بوجود آمدند که تغییرات عمده‌ی آن، جایگزین کردن لایه میانی RNN با یک بلاک،

[۳۰] $long\ short-term\ memory$



شکل ۴.۲: **شبکه‌های عصبی مصنوعی سیام:** (۱) لایه‌ی ورودی شبکه که شامل دو دسته ورودی مجزا است. (۲) لایه‌های مخفی شبکه با دو دسته‌ی مستقل ولی با وزن‌های یکسان در شبکه قرار گرفته‌اند. (۳) لایه‌ی شامل استخراج ویژگی برای سنجش میزان شباهت (۴) لایه‌ی بررسی میزان شباهت (۵) لایه‌ی خروجی، شامل امتیاز میزان شباهت دو ورودی.

که بلاک $LSTM$ نام دارد، است شکل ۳.۲.

بزرگترین ویژگی $LSTM$ امکان یادگیری وابستگی بلند مدت است که توسط شبکه های عصبی بازگشتی امکان پذیر نبود. برای پیش بینی گام زمانی بعدی نیاز است که مقادیر وزن‌ها در شبکه بروز رسانی شوند که این کار مستلزم حفظ اطلاعات گام‌های زمانی ابتدایی است. یک شبکه عصبی بازگشتی فقط می‌تواند تعداد محدودی از وابستگی‌های کوتاه مدت را یادبگیرد. سری‌های زمانی بلند مدت (به عنوان مثال هزار گام زمانی) قابل یادگیری توسط RNNها نیستند اما $LSTM$ها می‌توانند این وابستگی‌های بلند مدت را به درستی یاد بگیرند. [۴۱، ۴۲]

۶.۲.۲ شبکه‌های عصبی مصنوعی سیام

شبکه عصبی مصنوعی سیام [۳۱] ، یک شبکه عصبی مصنوعی است که از دو شبکه‌ی عصبی مشابه به صورت موازی استفاده می‌کند و دو بردار ورودی متفاوت، برای کشف بردارهای

[۳۱] $siamese\ artificial\ neural\ network$

خروجی قابل مقایسه، به عنوان ورودی، وارد شبکه می‌شوند. این شبکه‌ی عصبی مصنوعی برای پیدا کردن تصاویر مشابه (شناسایی امضاء) مورد استفاده قرار گرفت اما در ادامه برای مقایسه موارد مشابه در مجموعه‌های مختلف استفاده می‌شود. [۳۸، ۳۹]

این شبکه‌ی عصبی مصنوعی از ترکیب دو شبکه‌ی عصبی چندلایه تشکیل شده است، هر کدام از شبکه‌های مورد استفاده در شبکه‌ی عصبی مصنوعی سیام به یکی از طرفین مقایسه تعلق دارد. در شبکه‌ی عصبی سیام ورودی هر کدام از شبکه‌ها بطور موازی به شبکه بصورت مستقل داده می‌شود و پردازش می‌شود و خطای نتیجه پردازش قبل از خروجی گرفتن محاسبه می‌شود و به دو شبکه برای بروز شدن و یادگیری منتقل می‌شود شکل ۴.۲.

۷.۲.۲ نمایش ویژگی‌ها

قبل از بحث درباره ساختار شبکه‌های عصبی مصنوعی، توجه به نحوه نمایش ویژگی‌ها مهم است. در حال حاضر، ما می‌توانیم یک شبکه عصبی پیش‌خور پیش فرضی را به عنوان یک تابع $NN(x)$ که به عنوان ورودی، یک بردار x را می‌گیرد و یک بردار خروجی را تولید می‌کند، در نظر بگیریم. این تابع اغلب به عنوان یک دسته‌بند مورد استفاده قرار می‌گیرد. این تابع می‌تواند پیچیده باشد و تقریبا همیشه غیر خطی است. در اینجا، ما روی ورودی، x تمرکز می‌کنیم. در هنگام کار با زبان‌های طبیعی، ورودی x ویژگی‌هایی مانند: کلمات، بخشی از گفتار [۳۲] یا سایر اطلاعات زبان شناسی را رمزگذاری می‌کند. با استفاده از مدل‌های مبتنی بر شبکه‌های عصبی مصنوعی می‌توان هر ویژگی را به عنوان ابعاد منحصر به فرد به عنوان بردارهای متراکم، نمایش داد. بدین معنی که هر ویژگی اصلی به یک فضای چند بعدی، تعبیه شده و به عنوان یک بردار در آن فضای نمایش داده می‌شود.

شکل ۵.۲ دو رویکرد به نمایش ویژگی‌ها [۳۳] را نشان می‌دهد.جاسازی ویژگی [۳۴] به عنوان پارامترهای مدل، که باید با سایر اجزای شبکه آموزش دیده شوند (روش های آموزش جاسازی ویژگی ها در ادامه مورد بحث قرار می گیرند).

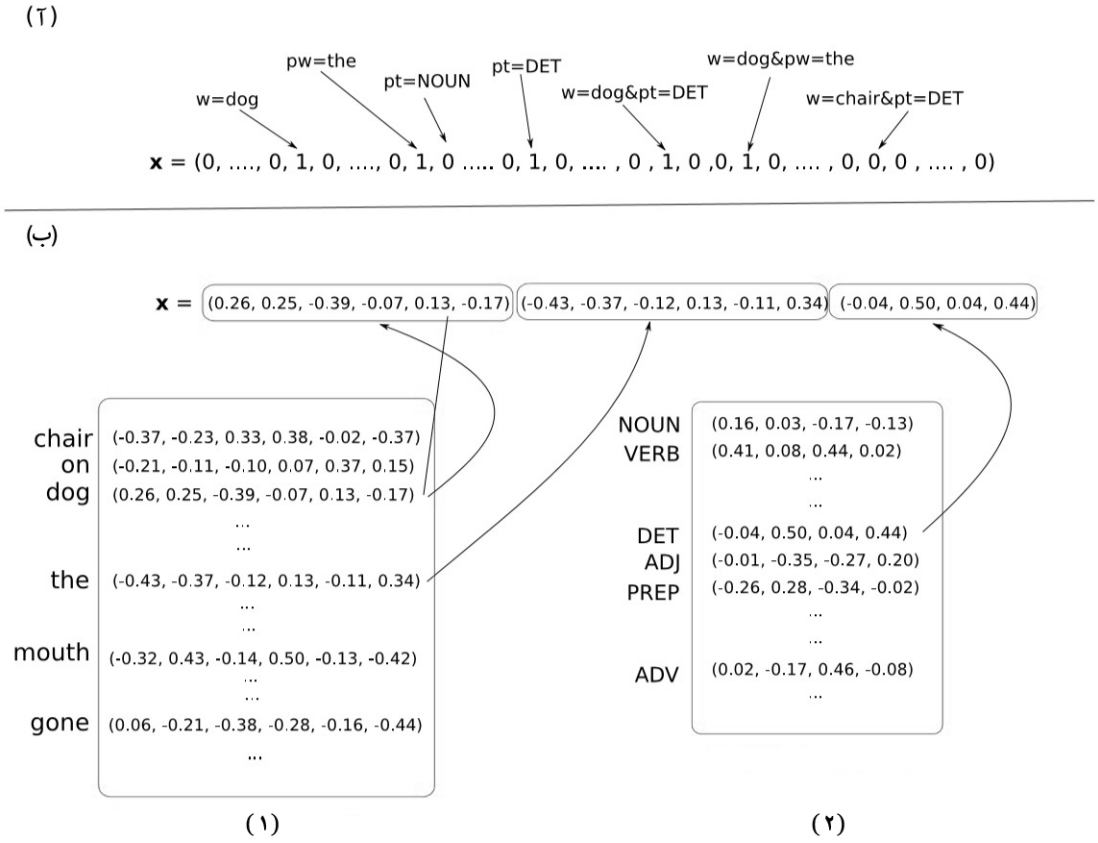
ساختار کلی سیستم طبقه بندی NLP بر مبنای شبکه عصبی مصنوعی پیش‌خور به این ترتیب است:

۱. مجموعه ای از ویژگی های اصلی زبانی [۳۵] $f_1 ... f_k$، که برای پیش بینی کلاس خروجی مناسب هستند را استخراج کنید.

۲. برای هر ویژگی f_i ، بردار مربوطه $v(f_i)$ آن را بدست آورید.

---

[۳۲] *part of speech*
[۳۳] *feature representation*
[۳۴] *feature embeddings*
[۳۵] *core linguistic features*



شکل ۵.۲: **نمایش بردار ویژگی تنک و بردار ویژگی متراکم** : (۱) جاسازی کلمه. (۲) جاسازی بخشی از گفتار. دو نمونه کدگذاری اطلاعات: کلمه‌ی مورد بحث "Dog" است و کلمه‌ی قبل آن "The" است و "DET" یک $pos - tag$ است. (آ) بردار ویژگی متغیر. هر بعد، نمایشگر یک ویژگی است و ترکیبی از ویژگی‌ها بردار را تشکیل می‌دهد. مقادیر ویژگی باینری هستند. ابعاد بسیار زیاد است. (ب) هر بردار ویژگی، مبتنی بر جاذبه چگال هر ویژگی اصلی ،به عنوان یک بردار نمایش داده می شود. هیچ رمزگذاری صریح ترکیبی از ویژگی‌ها وجود ندارد. ابعاد کم است. نگاشت ویژگی‌ها به یک بردار با اندازه مشخص، بردار هر کلمه را ایجاد می‌کند.

۳. بردارهای x ورودی را محاسبه کنید و تشکیل دهید (با توجه به روش‌های مختلف، می‌تواند این بردار از جمع کردن بردارهای ویژگی، ترکیب آن‌ها و یا هردو استفاده شود).

۴. بردار x را به عنوان ورودی شبکه عصبی مصنوعی، (شبکه عصبی مصنوعی پیش‌خور) را به شبکه وارد کنید.

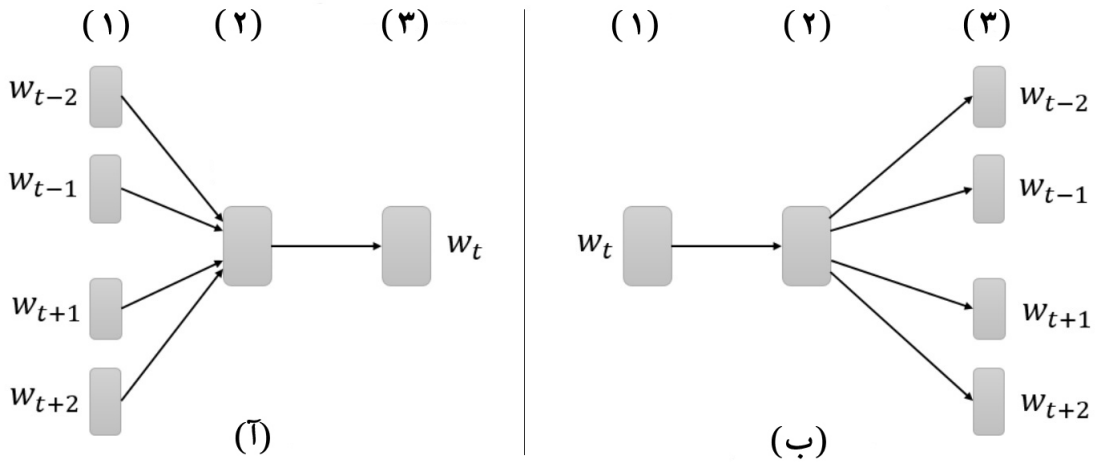
یکی از بزرگ‌ترین تغییری که در این دو دسته از نمایش ویژگی‌ها وجود دارد، ابعاد بزرگی است که از به هم پیوستن ویژگی‌ها در روش‌های تنک بوجود می‌آید و در مقابل در روش‌های متراکم ما ویژگی‌هایی داریم که می‌توانند باهم ترکیب شوند و در برداری با طول ثابت قرار گیرند. در ادامه جزئیات بیشتری از روش‌ها را توضیح می‌دهیم.

بردارهای متراکم در مقابل نمایش one-hot : مزایای نمایش ویژگی‌های متراکم، به عنوان بردارهایی با شناسه‌های منحصر به فرد، چیست؟ آیا همیشه لازم است ویژگی‌ها را با روشی که منجر به بردارهای متراکم می‌شود، ارائه کنیم؟

بگذارید دو نوع شیوه‌ی نمایش را در نظر بگیریم:

- ویژگی‌های روش نمایش $one-hot$:
  - ابعاد یک بردار $one-hot$ یکسان با تعداد ویژگی‌های متمایز آن است.
  - بردارها کاملا مستقل از یکدیگر هستند. به عنوان مثال بردارهای ویژگی کلمه‌ی "Dog" با بردار ویژگی کلمه‌ی "$Thinking$" و "Cat"، هیچ‌گونه شباهت و یا تفاوتی برنامه‌ریزی شده‌ای، ندارد .
- در مقابل ویژگی‌های روش نمایش متراکم را داریم که:
  - ابعاد بردارهای ویژگی ثابت (مقدار دلخواه) است.
  - ویژگی‌های مشابه بردارهای مشابهی دارند (اطلاعات بین ویژگی‌های مشابه به اشتراک گذاشته شده است).

یکی از مزایای استفاده از بردارهای متراکم، حجم محاسباتی کمتر نسبت به روش‌های تنک است. اکثر ابزارهای شبکه عصبی با بردارهای تنک، ناقص و ناکام هستند. ولی مزیت اصلی نمایش بردارهای متراکم در قدرت تعمیم آن است. اگر ما باور داشته باشیم برخی از ویژگی‌ها ممکن است سرنخ مشابهی را ارائه دهند، ارزشمند است که بردار نماینده‌ای را برای آن ارائه دهیم، که بتواند این شباهت‌ها را نمایش دهد. به عنوان مثال فرض کنید ما کلمه‌ی "dog" را چند بار در طول آموزش دیده‌ایم، اما کلمه‌ی "cat" را تنها چند لحظه مشاهده می‌کنید. اگر هر یک از کلمات دارای بردار مستقلی باشند، وقوع "dog" به ما هیچ چیز درباره وقوع "cat" نمی‌گوید. در بردارهای متراکم، بردار آموخته شده برای "dog" ممکن است شبیه به بردار آموخته شده‌ی "cat"، باشد. [۵، ۲۹، ۳۰]



شکل ۶.۲: **دو معماری مدلسازی روش Word2Vec** : (۱) ورودی شبکه. (۲) طرح ریزی. (۳) خروجی. (آ) معماری $CBOW$ (ب) معماری $Skip-Gram$

۸.۲.۲ جاسازی کلمات و مدل سازی احتمالاتی عصبی Word2Vec

جاسازی کلمات یکی از مراحل اصلی در پردازش زبان‌ها است. توانایی تبدیل متن به بردار بطوری که ، شباهت معنایی و نحوی، ارتباط با سایر کلمات و غیره نیز از آن قابل درک باشد.

$Word2vec$ یکی از محبوب ترین تکنیک‌های یادگیری جاسازی کلمات با استفاده از شبکه عصبی کم عمق است [۴۳]. در این روش سعی می‌شود با استفاده از شبکه عصبی کم عمق بردارهای متراکمی ساخته شود بطوری که بردار کلماتی که در اسناد زیاد در کنار هم (در یک عبارت، یک جمله، یک پاراگراف، یک متن) می‌آیند، در فضای چند بعدی عددی ساخته شده نیز به هم نزدیک باشند. $Word2Vec$ این عملیات را با دو نوع معماری کلی $Skip-gram$ و $CBOW$ [۳۶] انجام می‌دهد. در هر دوی این معماری ابتدا برای کلمات درون سند برداری با درایه‌های تصادفی تشکیل می‌دهیم. سپس با استفاده از شبکه عصبی بردار کلمات نزدیک به هم را به گونه‌ای تنظیم می‌کند تا مقادیر بردارهای نزدیک به هم، نزدیک شود شکل ۶.۲.

۹.۲.۲ مسائل پرسش_پاسخ

مسائل پرسش پاسخ (QA)، به عنوان مجموعه‌ای از مسائلی چون: خلاصه‌سازی و استخراج اطلاعات، جمع آوری داده‌ها، عبارات و یا پاراگراف‌ها و تحلیل و بررسی آن‌ها برای ایجاد یک ارتباط تعاملی بین انسان و رایانه است.

در حالی که QA می‌تواند به نحوی ترکیبی از این دو مسأله باشد اما مشکلاتی را در این شاخه‌ی پردازش زبان بوجود آورده، که تفاوت‌های بسیاری را ساخته و به راحتی مرزی را برای ایجاد یک علم جدید ایجاد نموده است. این مسئله از لحاظ تاریخی به زیر کارهایی چون:

[۳۶] $Continuous\ Bag\ of\ Words$

طبقه بندی سازی سوالات، بازیابی و استخراج پاسخ پرسش و تعیین نوع درخواست اطلاعات و فرمت پاسخ، تقسیم شده است. در بازیابی، خلاصه سازی و استخراج، ساده سازی و ترکیب اطلاعات در یک نظم هوشمند برای ایجاد پاسخ استفاده می‌شود. رویکردهای بسیار ابتدایی برای پاسخ دادن، به استخراج اطلاعات از پایگاه داده با روش‌های مختلف، تجزیه و تحلیل آن‌ها و ارائه‌ی عبارت بازیابی شده به عنوان یک پاسخ پیشنهادی، بر می‌گردد. [۵۶، ۳۲، ۳۳]

محققان حوزه‌ی پردازش_زبان‌های طبیعی در مقاله‌ی [۳۴] نشان داده‌اند، می‌توان از یک شبکه عصبی مبتنی بر توجه، استفاده کرد تا مسأله با یک پاسخ حاوی اطلاعات متناسب مطابقت کند. مکانیسم توجه، به طور خودکار به گونه‌ای مطابقت داده می‌شود تا از نمایندگی تمام پاراگراف استفاده کند. در نهایت، شبکه‌های اشارگر برای پیش‌بینی موقعیت و مرز یک پاسخ استفاده شدند. این شبکه‌ها با استفاده از نشان دادن بردارهای جمع آوری اطلاعات از پیام و همچنین کلمات مورد نظر تجزیه و تحلیل، برای مدل کردن نشانه‌های بحرانی یا عبارات ضروری برای درک لازم مطلب استفاده می کنند. این مکانیزم کمک کرد تا بتوان تمرکز را از کلمات معنی دار در سراسر گذرها به پنجره های متنی اطراف تغییر داد.

شبکه های عصبی کانولوشن چند ضلعی [۳۷] برای تجزیه و تحلیل خودکار سوالات از دیدگاه‌های مختلف ساخته شد، در ادامه از شبکه‌های موازی برای استخراج اطلاعات مربوطه از سوالات ورودی استفاده کردند. و شبکه‌های جداگانه را برای استخراج اطلاعات و روابط زمینه و برای مشخص کردن اینکه کدام فرم‌ها باید پاسخ داده شود، مورد استفاده قرار دادند. خروجی این شبکه‌ها سپس با یکدیگر ترکیب شد و برای رتبه‌بندی پاسخ‌های احتمالی استفاده شد. [۳۵]

یکی از رویکردهای اخیر، استفاده از شبکه های ارتباطی [۳۸] بود. RNها بر اساس یک معماری MLP ساخته می‌شوند، که در آن تمرکز بر استدلال‌های ارتباطی، یعنی روابط تعاملی بین شخصیت‌ها در داده‌ها است. این شبکه‌های پیش‌خور عملکرد مشابهی را در میان تمام جفت پرسش_پاسخ‌ها انجام می‌دهند تا همبستگی بین آن‌ها را به وجود آورند. ورودی RN ها، بردارهایی هستند که با استفاده از $LSTM$ از احکام درون سند ساخته می‌شوند. همه‌ی زیرمجموعه‌ها (احکام استخراج شده به وسیله‌ی $LSTM$) برای تعیین اینکه، آیا احکام استخراج شده در یک سند هست یا نه، یا بین این جملات و سوال مطرح شده است، وجود دارد. [۳۶، ۳۷]

۳.۲ مدل مفهومی

در این پایان‌نامه روشی مبتنی بر شبکه‌های عصبی مصنوعی مورد بررسی قرار گرفته می‌شود. در ابتدا با استفاده از شبکه‌ی عصبی به جاسازی کلمات می‌پردازیم و در ادامه با استفاده از پیدا

[۳۷] *multicolumn convolutional neural networks*

[۳۸] *relational networks*

کردن کلمات کلیدی درون پرسش‌ها و پاسخ‌ها پرداخته و با استفاده از تنظیم ضریب مناسب برای بردارهای کلمات کلیدی درون عبارات به ساختن بردار مناسب برای پرسش‌ها و پاسخ‌ها می‌پردازیم و در انتها با آموزش شبکه‌ی عصبی مصنوعی سیام سعی در ساختن یک سیستم پاسخگویی هوشمند می‌کنیم.

۴.۲ جمع بندی

همان‌گونه که مشاهده کردیم، پیشرفت‌های اخیر در تکنولوژی شبکه‌های عصبی مصنوعی، به ویژه یادگیری عمیق، به شدت در زمینه یادگیری ماشین تاثیر گذاشته است، به طوری که اغلب راه‌حل‌های مختلفی از پیچیده‌ترین مسائل در حوزه‌های گوناگونی را متوجه خود داشته است. پردازش زبان‌های طبیعی هم به عنوان یکی از مهم‌ترین شاخه‌های هوش مصنوعی در استفاده از این تکنولوژی مستثنا نبوده است.

در بسیاری از زمینه‌های NLP، استفاده از یادگیری عمیق نتایجی به دست داده است و از سایر روش‌های یادگیری ماشین و روش‌های آماری مورد استفاده در سال‌های قبل، پیشی گرفته است. در این قسمت به بررسی روش‌های رایج قبلی و جدید که در طول سال‌ها توسط دانشمندان مختلف ساخته شده است، پرداختیم.

فصل ۳

روش شناسایی تحقیق

۱.۳ مقدمه

سیستم پرسش_پاسخ خودکار یکی از مسائل کلاسیک در پردازش زبان‌های طبیعی است، که هدف آن طراحی سیستم‌هایی است که می‌توانند به طور خودکار به یک سوال پاسخ دهند، همان گونه که انسان انجام می دهد. ما یک مدل مبتنی بر یادگیری عمیق برای پاسخ به سوال به طور خودکار ارائه می‌دهیم.

در گام اول کلمات درون پایگاه داده با استفاده از یک مدل سازی احتمالات عصبی، جاسازی می‌شود. سپس سعی می‌شود بردار مربوط به پرسش‌ها و پاسخ‌ها، بر اساس بردارهای کلمات درون هر عبارت به گونه‌ای ساخته شود تا بردارهای سوالات و پاسخ‌ها بر اساس موضوعی که دارند، جاسازی شوند. سپس یک شبکه عصبی شباهت عمیق برای پیدا کردن نمره شباهت یک جفت پرسش و پاسخ آموزش داده می‌شود. و در انتها برای هر پرسش، بهترین پاسخ بر اساس بالاترین نمره همبستگی پیدا می‌شود. ما این مدل را در یک پایگاه داده عمومی در سطح وسیع آموزش می‌دهیم و سپس آن را برای استفاده در موضوعات مشخص، دقیقا تنظیم می‌کنیم.

۲.۳ روش و طرح کلی تحقیق

بسیاری از الگوریتم‌های کلاسیک برای پاسخ سوالات، شامل مراحل مختلفی نظیر: پردازش سوالات، طبقه بندی پاسخ و انتخاب پاسخ هستند و مهندسی زیادی در هر مرحله وجود دارد. [۴۶] بسیاری از الگوریتم‌های QA مدرن می‌آموزند که هر دو پرسش و پاسخ را به یک فضای چند بعدی نگاشت دهند و با پیدا کردن شباهت ویژگی‌هایشان پاسخ را انتخاب کنند. با توجه به عملکرد فوق‌العاده‌ای که شبکه‌های عصبی عمیق داشته‌اند، در سال‌های اخیر بسیاری از تحقیقات در این راستا بوده و از مدل‌های یادگیری عمیق برای طراحی سیستم پرسش‌ـ‌پاسخ استفاده کرده‌اند. در این کار ما یک مدل برای سیستم‌های پرسش‌ـ‌پاسخ با استفاده از شبکه‌ی شباهت عمیق پیشنهاد می‌کنیم. ابتدا یک مدل مبتنی بر شبکه عصبی برای جاسازی سوال و پاسخ با بردار نمایندگی چند بعدی استفاده می‌کنیم. سپس، این ویژگی‌ها به دو شبکه عصبی موازی تغذیه می‌شوند و پس از چند لایه‌ی متوالی، به منظور تصمیم گیری نهایی ترکیب می‌شوند.

۱.۲.۳ جاسازی پرسش‌ها و پاسخ‌ها

برای اینکه قادر به پردازش اسناد و داده‌های متنی با روش‌های یادگیری ماشین باشیم، نیاز داریم، یک نمایش برداری، برای آن‌ها ارائه دهیم. در بسیاری از روش‌های NLP، کلمات را به عنوان واحدهای اتمی داده‌های متنی در نظر می‌گیرند و هر کلمه را به عنوان یک شاخص در یک فرهنگ لغت واژگان نشان می‌دهند و بر اساس ویژگی‌های کیفی کلمه، برای پیدا کردن نمایندگی هر سند استفاده می‌کنند.

در ساده ترین حالت، در صورت استفاده از $unigram$، هر سند به عنوان یک هیستوگرام از کلمات موجود در آن نمایش داده می‌شود. با وجود سادگی این روش، این رویکرد برخی از محدودیت‌ها دارد. در این روش پردازش‌ها، بر اساس جایگاه کلمات، انجام می‌شود و اطلاعات معنایی واژگان نادیده گرفته می‌شود.

به همین دلیل در بسیاری از تحقیقات جدید، با استفاده از تکنیک‌های مبتنی بر شبکه عصبی سعی به یافتن بردار نمایش کلمه، می‌کنند به‌طوری که تشابه بین کلمات را نشان می‌دهند، یعنی کلمات مشابه، نمایندگی نسبتاً مشابهی در فضای تعبیه دارند. در اینجا شرح مختصری از برخی از طرح‌های تعبیه شده کلمات کلیدی با استفاده از شبکه عصبی را ارائه می‌دهیم و سپس آن را به مدل‌های مورد استفاده در این پایان‌نامه بسط می‌دهیم.

جاسازی کلمه، یک مسئله جدید در NLP نیست. [۴۴] یکی دیگر از کارهای رایج در این زمینه، [۴۵] که در آن تعبیه کلمه توسط یک شبکه عصبی با یک لایه‌ی پنهان صورت می‌گیرد. یک شبکه عصبی پیش‌خور را آموزش داده شد تا حداکثر احتمال پیش بینی یک کلمه را با

توجه به واژه‌های اطراف آن به حداکثر برسانند. ایده‌ی این کار بر این اساس است که؛ با توجه به شرایط هر کلمه (کلمات قبلی و بعدی)، تنها یک زیر مجموعه کوچک از کلمات وجود دارد که می‌تواند رخ دهد.

در این راستا روشی تحت عنوان $Word2Vec$ ساخته شد. [۴۷] دو مدل برای تولید یک بردار از کلمات را پیشنهاد شد، یکی از آنها به نام $continuous\ bag\ of\ words$ و دیگری $Skip-gram$ بود. در حالت $CBOW$، هدف یافتن نمایشی است، که با پیش‌بینی هر کلمه با استفاده از کلمات اطرافش تولید می‌شود. به عبارت دیگر، دنباله‌ای از کلمات $w۱...wN$، در دست داریم و هدف به حداکثر رساندن مقدار عددی زیر است:

$$\text{maximize } \frac{۱}{N}\sum_{i}^{N}\log p(w_i|w_{i-k},...,w_{i+k}) \qquad (۱.۳)$$

در حالی که در مدل $Skip-gram$ هدف یافتن بردار عددی برای کلماتی است که برای پیش بینی کلمات اطراف، در یک جمله یا یک سند مفید است [۴۸] . این روش حداکثر مقدار احتمال ورود را که کمی متفاوت از $CBOW$ محاسبه می کند:

$$\text{maximize } \frac{۱}{N}\sum_{i}^{N}\sum_{j=-k,j\neq ۰}^{k}\log p(w_{i+j}|w_i) \qquad (۲.۳)$$

این مدل‌ها معمولا با استفاده از رگرسیون $Softmax$ سلسله مراتبی [۱] و روش نمونه‌گیری منفی [۲] آموزش دیده‌اند. پس از آموزش این مدل‌ها، ما می‌توانیم از وزن لایه پنهان به عنوان جاسازی هر کلمه که توسط $F(.)$ مشخص شده است استفاده کنیم. با استفاده از این مدل، می‌توانید نمایش بسیار جالب از کلمات، که می‌تواند به نوعی مانند یک فضای برداری رفتار کنند را ارائه دهیم.

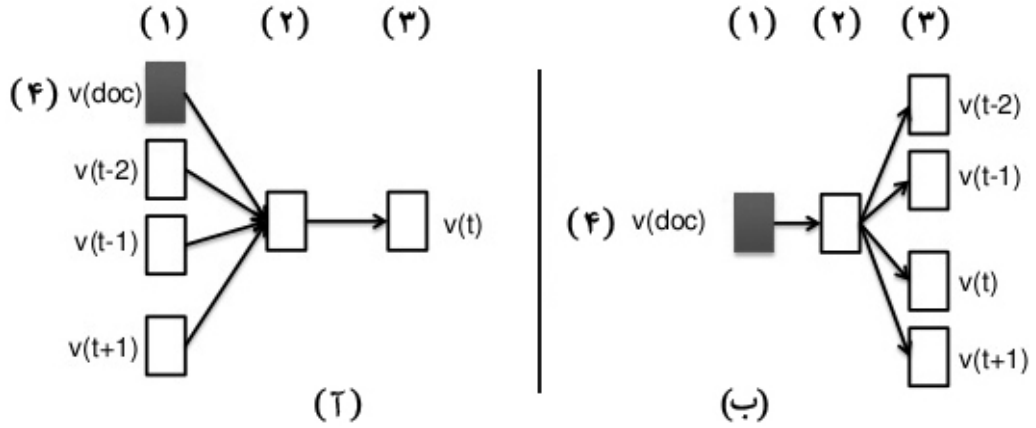
به عنوان مثال: محاسبه‌ی مقدار "$F(biggest)-F(big)+F(small)$" برداری است که نزدیک به بردار "$F(smallest)$" می‌باشد. که این امر نشان می‌دهد این روش قادر به یادگیری شباهت معنایی میان کلمات است.

در ادامه این ایده را برای یافتن یک بردار نمایش با طول ثابت برای متون با طول متغیر (که می تواند یک جمله، یک پاراگراف یا حتی یک سند باشد) گسترش داده‌اند. همانند مدل $Word2vec$ ، در مدل $Doc2vec$ یک بردار منحصربفرد در کنار ساختن بردار، برای بقیه‌ی کلمات ایجاد می‌شود که می‌تواند این بردار حاصل ترکیب یا مجموع (و ...) بردار کلمات مرتب با آن باشد شکل ۱.۳.

در این قسمت ما به جای استفاده از میانگین حسابی معمولی از میانگین حسابی وزنی استفاده می‌کنیم. در ابتدا کلمات کلیدی درون هر عبارت را محاسبه می‌کنیم، سپس بردار عبارت را از میانگین وزن دار کلمات درون عبارت محاسبه می‌کنیم. برای آن که بردار هر

[۱] $hierarchical\ Softmax\ regression$

[۲] $negative\ sampling\ technique$



شکل ۱.۳: **معماری Doc2Vec** : (۱) لایه‌ی ورودی (شامل بردار کلمات). (۲) لایه‌ی مخفی(بعلاوه ترکیب و یا میانگین گیری از بردار ایجاد شده برای کلمات، برای ساخت بردار سند). (۳) لایه‌ی خروجی(شامل بردار کلمات). (۴) بردار مربوط به سند . (آ) معماری DM. (ب) معماری $DBOW$.

عبارت بسته به موضوعی که دارد ساخته شود، ضریب کلمات کلیدی در هر عبارت را ضریب بالاتری انتخاب می‌کنیم تا بردار عبارت به کلمه‌ی کلیدی‌اش نزدیکتر شود. برای این منظور ابتدا با استفاده از کتابخانه‌ی $rake-nltk$ در زبان برنامه نویسی پایتون کلمات کلیدی در هر عبارت را بدست می‌آوریم. در ادامه بردارهای کلمات را نرمال می کنیم (به بازه‌ی باز(۰،۱) نگاشت می‌دهیم) و بسته به عبارت، کلمه یا کلمات کلیدی درون متن را با ضریب γ_i روی کلمات کلیدی اعمال می‌کنیم و سپس میانگین حسابی می‌گیریم. با انجام این کار باعث می‌شویم بردار پرسش‌ها و پاسخ‌ها علاوه بر این که به بردار کلمات درونشان بیشتر شبیه شوند، بردار پرسش با بردار پاسخ متناظرش نیز شبیه شود.

۲.۲.۳ شبکه‌های شباهت عمیق

پس از استخراج ویژگی‌ها، ما نیاز به آموزش مدلی داریم که یک پاسخ مناسب را در مقابل پرسش مطرح شده ارئه دهد. راه‌های مختلفی برای دستیابی به این هدف وجود دارد. در یک روش بسیار ساده می‌توان از ویژگی‌های $Doc2vec$ ساخته شده برای پرسش‌ها و پاسخ‌ها استفاده کرد و یک طبقه بند را که احتمال تطبیق را پیش‌بینی می‌کند، آموزش داد. در این کار، از شبکه‌های عصبی مصنوعی سیام [۴۹، ۵۰] استفاده کرده‌ایم و بردارهای پرسش و پاسخی را که ساختیم به عنوان ورودی این شبکه‌ی عصبی به شبکه ارسال می‌کنیم. آنگاه این بردارها بعد از چند لایه پردازش، بردار ویژگی‌هایی را برای سنجش دو عبارت تشکیل می‌دهند، در انتها این بردارها با یکدیگر تطبیق داده می شوند و در صورت اختلاف میزان خطا به لایه‌های قبل برگشت داده می‌شود تا وزن‌های لایه‌های شبکه‌ی عصبی به روز شود، این عملیات در فاز

آموزش تا آنجا ادامه پیدا می‌کند تا شبکه بتواند قدرت بررسی میزان شباهت را بدست آورد. در انتها از این شبکه برای سنجش میزان شباهت یک جفت پرسش_پاسخ استفاده می‌شود.

باید توجه داشته باشیم که در شبکه اصلی سیام، دو شبکه محدود با وزن‌های یکسان قرار گرفته اند، در حالی که در مدل ما چنین محدودیتی وجود ندارد. دلیل آن این است که شبکه اصلی سیام برای پیدا کردن شباهت چهره (و امضا در تصاویر [۵۱]) پیشنهاد شده است، بنابراین هر دو ورودی آن چهره هستند، و منطقی است که یک شبکه مشابه داشته باشد که نمایه‌ی انتزاعی خود را پیدا کند. در حالی که در مورد ما یکی از شبکه‌ها، سوال را پردازش می کند و دیگری پاسخ را پردازش می‌کند که واضح است که به وزن‌های مختلف نیاز داشته باشند تا پرسش و پاسخ را به یک فضای مشترک نگاشت دهند.

۳.۳ پایگاه داده

تحقیقات و آزمایشات این پایان نامه بر روی پایگاه داده‌ی، مجموعه‌ی پرسش_پاسخ‌های مربوط به شرکت بیمه صورت گرفته است که در ادامه به شرح کامل آن می‌پردازیم. این پایگاه داده حاوی پرسش‌ها و پاسخ‌های جمع آوری شده از وب سایت کتابخانه‌ی شرکت بیمه [۳] است. محتوای این پایگاه داده شامل پرسش‌های کاربران دنیای واقعی است و پاسخ‌های با کیفیت بالا توسط متخصصین با دانش عمیق، ساخته شده است. این پایگاه داده شامل سه مجموعه به صورت؛ داده‌های فاز آموزش و دادهای فاز اعتبارسنجی و داده‌های فاز آزمون می‌باشد که در جدول ۱.۳ به نمایش گذاشته شده است. در مقابل هر پرسش بیش از یک پاسخ وجود دارد.نمونه‌ای از یک پرسش_پاسخ در شکل ۲.۳ نمایش داده شده است.

۴.۳ جمع‌بندی

در این فصل به تفصیل به در رابطه با روش پیشنهادی‌مان در رابطه با بهبود سیستم‌های پرسش_پاسخ با استفاده از روش جدیدی بر پایه‌ی *Doc2vec* ساختیم و با استفاده‌ی صحیح از شبکه‌های عصبی سیام و کمی تغییر در روند آموزش این شبکه توانستیم این شبکه‌ی عصبی عمیق را سریع‌تر و دقیق‌تر برای بررسی میزان شباهت یک جفت پرسش_پاسخ استفاده کنیم.

[۳] http://www.insurancelibrary.com/

جدول ۱.۳: مشخصات پایگاه داده

	پرسش‌ها	پاسخ‌ها
داده‌های آموزش	۱۲۸۸۹	۲۱۳۲۵
داده‌های اعتبارسنجی	۲۰۰۰	۳۳۵۴
داده‌های آزمون	۲۰۰۰	۳۳۰۸
مجموع	۱۶۸۸۹	۲۷۹۸۷

What Is A Health Insurance Claim Form 1500?

Until recently when you heard the term "self funded health insurance" you could have safely assumed that you were talking about an employer sponsoredgroup health insurance plan funded either entirely through employer/employee dollars or a combination of these with some re-insurance policies included. In self funding the employer is taking on the risk for health care expenses either in full or in part (if re-insurance is involved). Under fully insured arrangements the employer has no further financial risk after the monthly premium has been paid. The has been a tendency lately by some in the media and in government to use the term self fund or self pay interchangeably in relation to the purchase and ownershipof individual health insurance policies. By definition, even after the ACA, virtually everyone in the individual health insurance marketplace is a self payor by virtue of the monthly premium they pay but they are not self funded because they have passed off all risk after any plan deductibles, copays and co-insurance to their insurance carrier in exchange for their monthly premium.

شکل ۲.۳: **نمونه‌ای از پرسش‌ـ پاسخ درون پایگاه داده**

فصل ۴

نتیجه‌گیری

۱.۴ مقدمه

در فصل قبل مشاهده کردیم که با استفاده از شبکه‌های عصبی عمیق می‌توانیم پرسش‌ها و پاسخ‌ها را در یک فضای با خصوصیات مشابه مدل کنیم و همچنین در بین روش‌های موجود، استفاده از روش‌های مبتنی بر مدل‌سازی احتمالاتی عصبی منجر به کامل شدن این مقصود می‌شوند در ادمه با آموزش شبکه‌ی عصبی عمیقی که به منظور پیدا کردن ویژگی‌های مشابه جفت پرسش_ پاسخ تنظیم شده است، آموزش داده می‌شود و در انتها پرسش‌ها و پاسخ‌های متناظر ما را به سمت یک سیستم پرسش_ پاسخ می‌رساند تا با استفاده از این روش بتوانیم برای پرسش‌های مطرح شده توسط پرسشگر برای یک سیستم پرسش_پاسخ، به دنبال پاسخ صحیح در پایگاه داده بگردیم و پاسخ مناسب را ارائه دهیم.

در ادامه به بررسی قدرت و ضعف این روش و مقایسه‌ی آن با روش‌های مشابه می‌پردازیم. و معیارهای مختلف را اندازگیری می‌کنیم و در ادامه در رابطه با محدودیت‌های این روش و کارهایی که می‌تواند منجر به بهبود این سیستم شود، بحث می‌کنیم.

۲.۴ پیشنیاز

۱.۲.۴ معرفی چند معیار مقایسه

در مباحث علم داده لازم است با معیارهای ارزیابی سنجش الگوریتمهای این حوزه آشنا شویم. در اغلب موارد ما به دنبال ساخت یک مدل بر اساس داده های موجود برای پیش‌بینی مقدار یک ویژگی از آن‌ها هستیم. چه این ویژگی یک مقدار غیر عددی باشد مانند تعیین مشروط شدن یا نشدن دانشجو، ترک کار شرکت توسط کارمند، تعیین موضوع یک توئیت و مانند آن که به آن‌ها مسائل دسته‌بندی میگوئیم و یا یک مقدار عددی مانند حدس زدن میزان حقوق یک فرد، قیمت یک خانه و مانند آن که به آن‌ها مسائل درونیابی یا رگرسیون می گوییم.

حتی در مسایل و الگوریتمهای بدون ناظر، مانند خوشه بندی نیز به معیارهایی برای تعیین دقت مدل نیاز خواهیم داشت. بنابراین در ادامه، معیارهای اصلی مورد استفاده در ارزیابی الگوریتمها و مدلهای مختلف علم داده را مرور خواهیم کرد.

۲.۲.۴ نگاهی دقیق‌تر به مسائل دسته‌بندی

در مسائل دسته‌بندی به دنبال پیش‌بینی دسته یا گروهی هستیم که یک رکورد یا داده به آن تعلق دارد. مثلا می‌خواهیم با داشتن مشخصات مختلف یک دانشجو مانند رشته، سهمیه قبولی، وضعیت مالی خانواده، تعداد هم اتاقی‌ها ، وضعیت تحصیلی دوران دبیرستان، تعداد واحدهای گذرانده، معدل ترم‌های پیش، تعداد ترم‌های مشروطی و مانند آن تعیین کنیم که این دانشجو مشروط خواهد شد یا نه . یا به عبارتی آیا به دسته مشروطی‌ها تعلق دارد؟

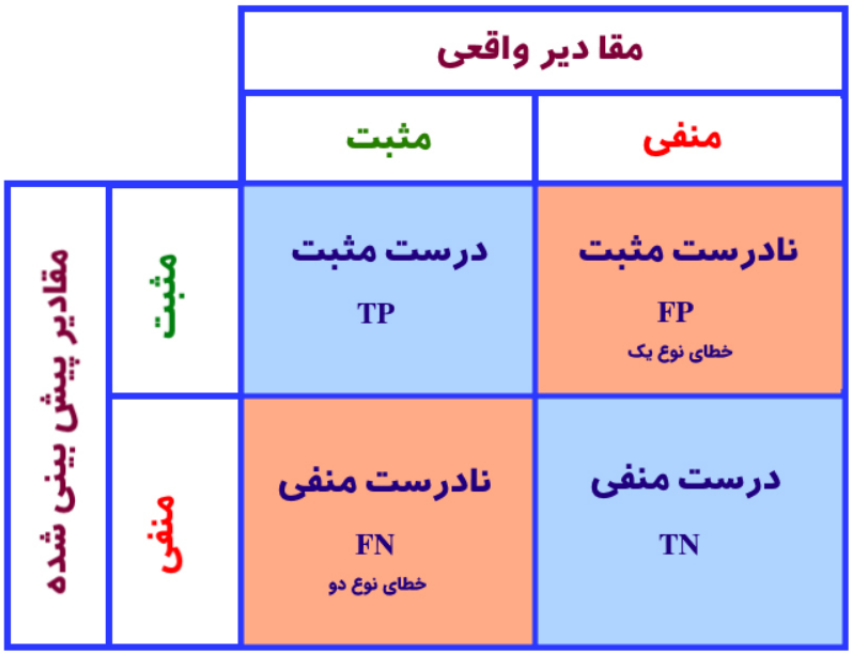
در واقع دسته‌بندیهای دوگانه بیشترین کاربرد را در دنیای واقعی دارند و ما هم برای ساده کردن مطلب، تمرکز را بر این نوع دسته‌بندی خواهیم گذاشت اما تمامی موارد بیان شده، به راحتی قابل تعمیم به دسته‌بندی های چندگانه هم خواهند بود.

بعد از ساخت یک مدل دسته‌بندی یعنی یافتن الگوریتمی که با مشاهده یک داده جدید، دسته‌ی آن را مشخص کند، برای سنجش میزان کارآیی و دقت مدل پیشنهادی، آن را بر روی داده‌های آموزشی یعنی داده‌هایی که از قبل دسته‌بندی آن‌ها را میدانیم، اعمال میکنیم.

خروجی این مدل در شکل ۱.۴ و ۲.۴ به تصویر کشیده شده است. با فرض اینکه هدف ما پیش‌بینی دیابت یک بیمار باشد یعنی اگر پیش‌بینی مثبت باشد یعنی بیمار، مبتلا به دیابت است و اگر پیش بینی منفی باشد، یعنی بیمار به دیابت مبتلا نیست، به تحلیل سلول‌های این ماتریس می‌پردازیم :

درست مثبت (TP) : اگر بیمار واقعا دیابت داشته باشد و مقدار پیش‌بینی شده هم دیابت را نشان دهد.

نادرست مثبت (FP) : اگر بیمار دیابت نداشته باشد اما نتیجه پیش بینی ما، نشانگر دیابت



شکل ۱.۴: **ماتریس درهم‌ریختگی یا اغتشاش**

بیمار باشد.

نادرست منفی (FN) : اگر بیمار دیابت داشته باشد اما پیش بینی ما، دیابت را منفی نشان دهد.

درست منفی (TN) : اگر بیمار دیابت نداشته باشد و پیش بینی ما هم همین را نشان بدهد.

حال آن‌که با یک سری پیش‌نیازها آشنا شدیم به تعریف اصطلاحات زیر می‌پردازیم.

۳.۲.۴ Accuracy

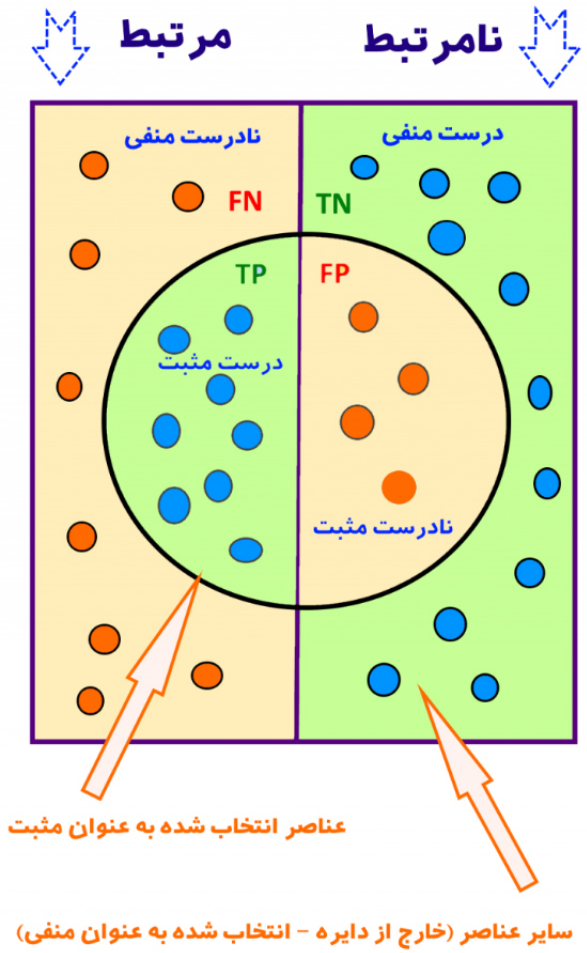
این معیار بر اساس داده‌هایی که انتخاب شده‌اند و داده‌هایی که انتخاب نشده‌اند ساخته می‌شود و می‌تواند دید نسبتا خوبی در برخی کاربردها را ارائه دهد.

$$Accuracy = (TP + TN)/(TN + FN + FP + TP) \qquad (۱.۴)$$

۴.۲.۴ Precision

یکی دیگر از معیارهای ارزیابی بسیار مهم و پر استفاده در علم داده‌کاوی، معیار صحت [۱] می‌باشد. این برای بررسی و پاسخ به این سوال ساخته می‌شود که، آیا الگوریتم کارایی دارد و اصطلاحا به صرفه است برای استفاده کردن؟

[۱] precision



شکل ۲.۴: **نمایش بصری ماتریس درهم‌ریختگی**

$$Precision = TP/(TP + FP) \quad (۲.۴)$$

Recall ۵.۲.۴

معیار پرکاربرد دیگری با نام بازخوانی [۲] وجود دارد که بسیار در داده‌کاوی از این فاکتور ارزیابی استفاده می‌شود. در این فاکتور تعداد نمونه‌های مثبت شناسایی شده مورد تحلیل و بحث قرار می‌گیرد.

$$Recall = TP/(TP + FN) \quad (۳.۴)$$

$F_۱$ ۶.۲.۴

در انتها به معرفی معیار $F_۱$ می‌پردازیم، این معیار بر اساس دو معیار مهم $precision$ و $recall$ ساخته می‌شود و روش بدست آوردن آن به شرح زیر است.

$$F_۱ = ۲ * ((precision * recall)/(precision + recall)) \quad (۴.۴)$$

۳.۴ خلاصه تحقیق

در این جا ما آزمایشات‌مان را بر اساس پایگاه داده‌ای که در دسترس داشتیم و همچنین نتایج برخی از کارهای مشابه می‌سنجینم و نتایج را ارائه می‌دهیم و همچنین این روش را از برخی جنبه‌هایی که به نظر می‌رسد بتواند در بهبود الگوریتم و در نهایت بالا بردن کارایی روش، موثر باشد را مورد بحث و آزمایش قرار می‌دهیم.

۴.۴ بحث در تحقیق

این قسمت را به چند بخش تقسیم می‌کنیم و نتایج برخی از تغییرات پیشنهادی را بر روی سیستم پرسش‌ـ پاسخ‌مان به نمایش می‌گذاریم.

---

[۲] recall

۱.۴.۴ جاسازی عبارات

در این قسمت نتایج بدست آمده از روش‌های مختلف جاسازی عبارات و تاثیر آن بر عملکرد شبکه‌ی عصبی سیام را مورد بررسی قرار می‌دهیم. در پایگاه داده‌ای که در دسترس داریم تعداد ۲۰۰۰ پرسش با ۳۳۰۸ پاسخ برایشان مشخص شده است. برای آنکه بتوانیم راحت‌تر پاسخ پرسش‌های مطرح شده را پیدا کنیم، سعی کردیم بردارهای پرسش‌ها و پاسخ‌ها را که عباراتی هستند که از کلمات تشکیل شده‌اند را از روی بردارهای کلمات بدست آوریم. برای این کار ما از دو دسته از روش‌ها استفاده کرده‌ایم، اول روش *doc2vec* و روش‌های دیگر استفاده از روش‌های میانگین گیری مختلف می باشد.

در ابتدا بردارهای پرسش‌ها و پاسخ‌ها را بر اساس میانگین‌گیری از بردار کلمات درون آن‌ها، با روش‌های میانگین‌گیری متفاوت ساختیم و برای مشخص کردن میزان شباهت بین پرسش‌ها و پاسخ‌ها، بردارهای حاصل را به الگوریتم نهایی داده‌ایم تا مشخص کنیم در کدام روش می‌توانیم تعداد جفت‌های صحیح بیشتری ایجاد کنیم. نتایج این آزمایش در جدول ۱.۴ و جدول ۲.۴ نشان داده شده است و از هیستوگرام ۳.۴ برای شفافیت بیشتر استفاده شده است.

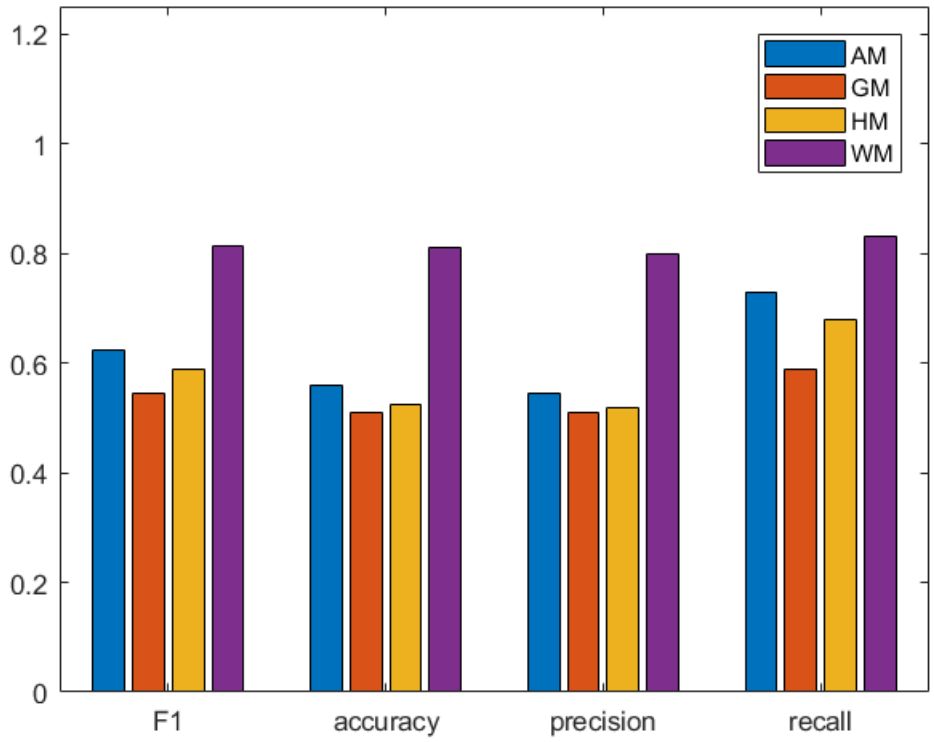
جدول ۱.۴: بررسی روش‌های مختلف میانگین گیری در ساخت بردار عبارات بر اساس میانگین گیری از بردار کلمات درون عبارات.

روش‌ها	TP	FN	TN	FP
میانگین حسابی	۲۴۱۴	۸۹۴	۱۲۹۰	۲۰۱۸
میانگین هندسی	۱۹۵۱	۱۳۵۷	۱۴۲۲	۱۸۸۶
میانگین هارمونیک	۲۲۴۹	۱۰۵۹	۱۲۲۳	۲۰۸۵
میانگین وزنی	۲۷۴۵	۵۶۳	۲۶۱۳	۶۹۵

همانگونه که در جدول ها مشخص شده است با استفاده از روش‌های مختلف میانگین‌گیری می‌توان بردار عبارات پرسش و پاسخ را ساخت. همان طور که دیده می‌شود نتایج حاصل از استفاده از روش میانگین وزنی نتایج بهتری را ارائه می‌دهد و دلیل آن مربوط می‌شود به این مساله که با استفاده از این مدل می‌توان بردارهای جفت پرسش‌ـپاسخ را به یکدیگر نزدیک کرد به این صورت که با وزن دهی به کلمات کلیدی درون هر عبارت و میانگین گیری سعی

جدول ۲.۴: محاسبه معیارهای مقایسه برای استفاده از میانگین‌گیری‌های متفاوت برای ساخت بردار پرسش‌ها و پاسخ‌ها.

روش‌ها	accuracy	precision	recall	F_1
میانگین حسابی	۰.۵۵۹۸	۰.۵۴۴۶	۰.۷۲۹۷	۰.۶۲۳۷
میانگین هندسی	۰.۵۰۹۸	۰.۵۰۸۴	۰.۵۸۹۷	۰.۵۴۶۰
میانگین هارمونیک	۰.۵۲۴۸	۰.۵۱۸۹	۰.۶۷۹۸	۰.۵۸۸۵
میانگین وزنی	۰.۸۰۹۸	۰.۷۹۷۹	۰.۸۲۹۸	۰.۸۱۳۵



شکل ۳.۴: **مقایسه‌ی چند روش میانگین‌گیری برای ساخت بردار عبارات پرسش و پاسخ از لحاظ معیارهای مختلف :** AM) میانگین حسابی، GM) میانگین هندسی، HM) میانگین هارمونیک، WM) میانگین وزنی.

می‌کنیم عباراتی را که از لحاظ معنایی به یکدیگر نزدیک می‌باشند را از لحاظ برداری نزدیک به هم قرار دهیم.

حال آن‌که روش‌های مختلف میانگین‌گیری را مقایسه کردیم و به این نتیجه رسیدیم که روش میانگین‌گیری وزنی می‌تواند عملکرد بهتری را در ساخت بردار عبارات رائه دهد. در ادامه میزان شباهتی را که این روش‌ها در بردار پرسش با بردار پاسخ ایجاد می‌کند را نسبت به بردارهای $doc2vec$ می‌سنجیم. این سنجش اینگونه است که، ابتدا بردارهای پرسش و پاسخ را (برای هر روش) می‌سازیم و بعد از نرمال کردن تمام بردارها در بازه‌ی (۰،۱) میانگین فاصله‌ی تمام جفت پرسش‌ـ پاسخ‌ها را محاسبه می‌کنیم. در آخر برای مقایسه مقادیر بدست آمده را بر عدد حاصل بر اساس روش $doc2vec$ تقسیم می‌کنیم که نتایج آن در جدول ۳.۴ نمایش داده شده است .

جدول ۳.۴: نسبت میانگین فاصله‌ی جفت پرسش‌ـپاسخ‌ها بر اساس بردار عبارات به روش‌های مختلف میانگین گیری کلمات با روش $doc2vec$.

روش‌ها	تغیرات میزان شباهت نسبت به روش $doc2vec$
میانگین حسابی	۰.۸۷۴۷
میانگین هندسی	۰.۸۰۸۲
میانگین هارمونیک	۰.۶۲۰۴
میانگین وزنی	۱.۲۰۸۶

در ابتدای کار، زمانی که به این نتیجه رسیدیم که برای آن که شباهت جفت پرسش‌ـ پاسخ را بالا ببریم، می‌توانیم کلمات کلیدی در هر عبارت پیدا کرده و طوری بردار عبارات ساخته شود تا شباهت عبارات متناظر (پرسش و پاسخش) بالا رود. بهترین کار برای این که این عملیات محقق شود، استفاده از میانگین‌گیری وزنی می‌باشد. بعد از آنکه به این نتیجه رسیدیم که از میانگین‌وزنی استفاده کنیم، حال مسأله‌ی دیگری مطرح می‌شود، این که این وزن دهی چگونه و بر چه اساس صورت گیرد.

بدین منظور روش‌های مختلفی را برا وزن دهی آزمایش کردیم که در نهایت می‌توان این روش‌ها را به دو صورت وزن‌های ثابت و وزن‌های متغیر دسته بندی می‌کنیم.

در روش وزن دهی ثابت تمام بردارهای کلمات کلیدی هر عبارت را در یک عدد ثابت ضرب می‌کنیم ولی در روش متغیر سعی می‌کنیم برای هر کلمه‌ی کلیدی در عبارت ضریب را انتخاب

کنیم به طوری که میزان شباهت بین جفت پرسش_پاسخ را تا حد ممکن بالا ببریم.

در این روش، کلمات کلیدی هر عبارت استخراج می‌شود آنگاه بر اساس اولویت‌بندی کلمات کلیدی مهم‌تر ضریب بالاتری را می‌گیرند. در این صورت بردار هر عبارت به بردار کلمه‌ی کلیدی‌اش، که در واقع به نحوی عبارت حول همان کلمه می‌چرخد، بسیار نزدیک می‌شود. مقایسه‌ی این دو روش نسبت به روش $doc2vec$ در جدول ۴.۴ نمایش داده شده است.

جدول ۴.۴: بررسی میزان شباهت در میانگین گیری وزنی پویا و میانگین‌گیری وزنی ساده نسبت به روش $doc2vec$

روش‌ها	تغیرات میزان شباهت نسبت به روش
میانگین وزنی با وزن ثابت	۱.۰۵۵۸
میانگین وزنی با وزن متغیر	۱.۲۰۸۶

پس از پیدا کردن روشی برای جاسازی مناسب جفت پرسش_پاسخ، سعی می‌کنیم این روش جاسازی را با روش $doc2vec$ مقایسه می‌کنیم و تاثیر آن را بر روی الگوریتم سیام در جدول ۵.۴ و ۶.۴ آورده ایم و برای وضوح بیشتر از هیستوگرام شکل ۴.۴ استفاده شده است.

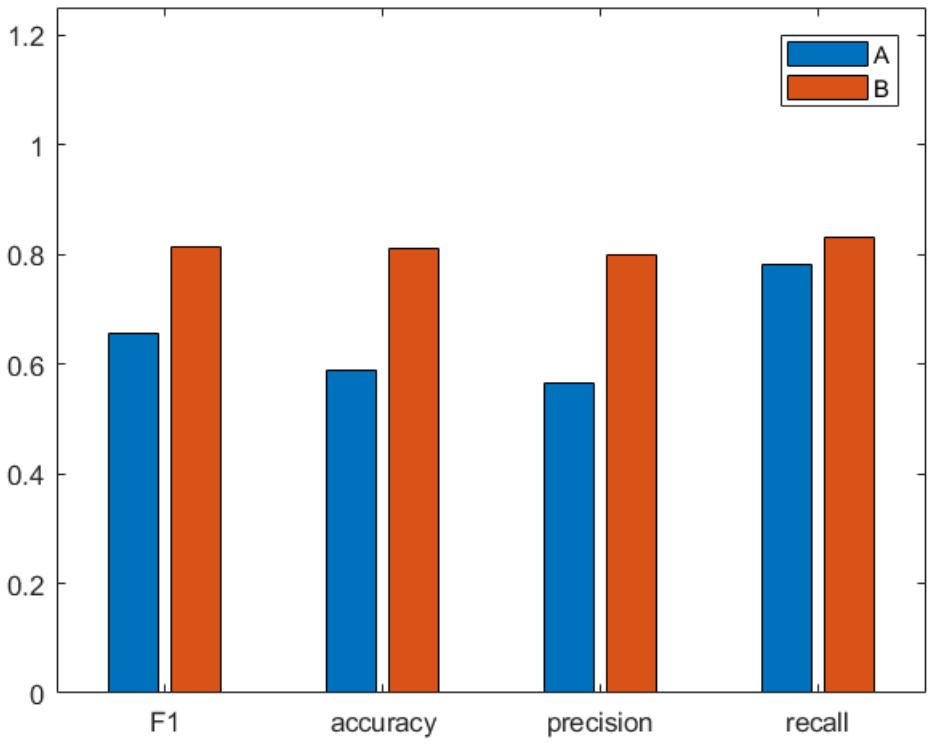
جدول ۵.۴: بررسی تعداد پاسخ صحیح به سوالات مطرح شده (تعداد کل پرسش‌های مرحله‌ی آزمون ۲۰۰۰ پرسش و مجموع پاسخ‌های مرحله‌ی آزمون ۳۳۰۸ پاسخ می‌باشد).

روش‌ها	TP	FN	TN	FP
$SNN + doc۲vec$	۲۵۸۰	۷۲۸	۱۳۲۳	۱۹۸۵
روش جاسازی پیشنهادی + SNN	۲۷۴۵	۵۶۳	۲۶۱۳	۶۹۵

حال آنکه روش جاسازی مدنظرمان انتخاب شد، به بررسی قسمت‌های دیگر سیستم پرسش_ پاسخ می‌رویم و سعی می‌کنیم با بهبود آن، سیستم پرسش_پاسخ قوی‌تری را بسازیم.

جدول ۶.۴: محاسبه معیارهای مقایسه در دو روش مختلف برای ساخت بردار عبارات .

روش‌ها	accuracy	precision	recall	F_1
$SNN + doc2vec$	۰.۵۸۹۹	۰.۵۶۵۱	۰.۷۷۹۹	۰.۶۵۵۳
روش جاسازی پیشنهادی + SNN	۰.۸۰۹۸	۰.۷۹۷۹	۰.۸۲۹۸	۰.۸۱۳۵



شکل ۴.۴: **مقایسه‌ی دو روش ساخت سیستم پرسش پاسخ اتوماتیک از لحاظ معیارهای مختلف:** (A $SNN + doc2vec$، B روش جاسازی پیشنهادی + SNN).

۲.۴.۴ شبکه عصبی سیام

شبکه‌ی عصبی سیام در پردازش تصویر و برای شناسایی صحت امضاء به کار می‌رود. در واقع ایده‌ی ساخت شبکه‌ی عصبی سیام به مقایسه‌ی دو شئ و پیدا کردن ویژگی‌هایی از آن‌ها برای سنجش میزان شباهت‌شان مربوط می‌شود. از آنجایی که دو شئ‌ای که با یکدیگر مقایسه می‌شوند، شباهت‌های ساختاری دارند، به عنوان مثال در مقایسه‌ی دو چهره، هر دو چهره دارای ساختار مشابهی می‌باشند، در نتیجه می‌توان چاچوب یکسانی را برای دو شئ مورد سنجش اتخاذ کرد. اما در پردازش متن و درواقع پیدا کردن پاسخ برای پرسش‌ها ما برای پرسش و پاسخ متناظرش چنین چارچوبی را در اختیار نداریم.

همانگونه که پیش‌تر بیان شد، شبکه‌ی عصبی سیام از دو شبکه‌ی عصبی موازی تشکیل شده است و هر قسمت مربوط می‌شود به اشیایی که قرار است با یکدیگر مقایسه شوند. در سیستم پرسش‌ـ‌پاسخ قسمتی از این شبکه مربوط است به پرسش‌ها و قسمت دیگر به پاسخ های متناظر با پرسش‌ها مربوط می‌شود. در هر قسمت ویژگی‌های هر شئ استخراج می‌شود، تا در ادامه‌ی شبکه‌ی عصبی، این دو شئ با یکدیگر مقایسه شوند. در فاز آموزش، این شبکه‌ی عصبی در جریان آموزش و مقایسه‌ی دو شئ، به روش انتشار به عقب، وزن لایه‌های شبکه‌ی عصبی را به روزرسانی می‌کند و این در حالی است که وزن‌های هر دو شبکه‌ی عصبی بصورت یکسان تغییر می‌کنند.

با توجه به آن‌که عبارات پرسش و پاسخ، چارچوب ساختاری متفاوتی دارند، لذا برای هر شبکه بصورت مستقل عملیات به روز رسانی صورت می‌گیرد. در جدول ۷.۴ و ۸.۴ و شکل ۵.۴ عملکرد دو شبکه‌ی عصبی سیام (به صورت استاندارد و بصورت تغییر یافته) مورد بحث قرار گرفته است.

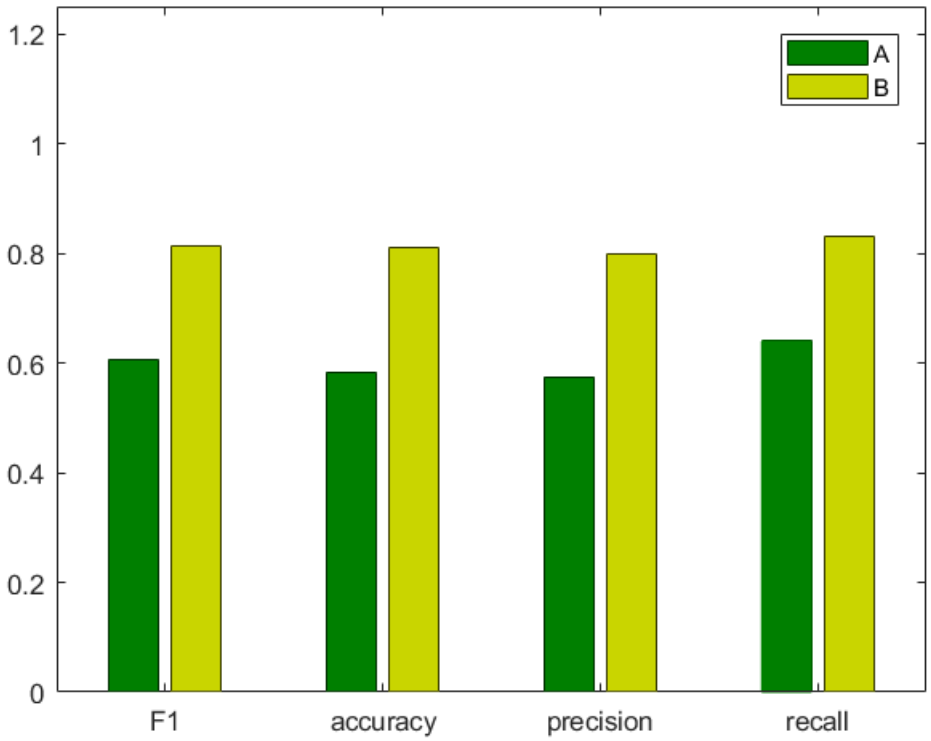
جدول ۷.۴: مقایسه عملکرد دو شبکه‌ی عصبی سیام بصورت شبکه‌ی عصبی سیام استاندارد و شبکه‌ی عصبی سیام تغییر یافته در تعداد تکرار (فاز آموزش شبکه‌ی عصبی) ۶۰۰ دور.

روش‌ها	TP	FN	TN	FP
شبکه‌ی عصبی سیام استاندارد	۲۱۱۸	۱۱۹۰	۱۷۳۰	۱۵۷۸
شبکه‌ی عصبی سیام با وزن‌های مستقل	۲۷۴۵	۵۶۳	۲۶۱۳	۶۹۵

با توجه به بررسی روش‌هایی که می‌تواند در این کلاس از پیاده سازی الگوریتم‌های که منجر به ساخت سیستم پرسش و پاسخ می‌شود، در نهایت به دو مقایسه‌ی دو روشی می‌پردازیم که از تمامی جهت‌های تاثیر گذار بررسی شده‌اند و نتایج بهتری را نسبت به روش‌های مقابل دارند.

جدول ۸.۴: محاسبه معیارهای مقایسه در دو روش مختلف استفاده از شبکه‌ی عصبی سیام. .

روش‌ها	accuracy	precision	recall	F_1
شبکه‌ی عصبی سیام استاندارد	۰.۵۸۱۶	۰.۵۷۳۰	۰.۶۴۰۲	۰.۶۰۴۷
شبکه‌ی عصبی سیام با وزن‌های مستقل	۰.۸۰۹۸	۰.۷۹۷۹	۰.۸۲۹۸	۰.۸۱۳۵



شکل ۵.۴: **مقایسه‌ی دو روش ساخت سیستم پرسش پاسخ اتوماتیک بر اساس نوع شبکه عصبی سیام:** (A شبکه عصبی سیام استاندارد (B شبکه عصبی سیام با وزن‌های مستقل.

در انتها روشی که دکتر مینایی ارائه کرده‌اند و روش پیشنهادی ما با یکدیگر مقایسه شده است و همانگونه که در جدول ۹.۴ نشان داده شده است روش ما به میزان ۲۰ درصد خطای عملکرد را بهبود داده است.

جدول ۹.۴: مقایسه روش مقاله‌ی دکتر مینایی و روش پیشنهادی.

روش‌ها	accuracy
$S.Minaee - ۲۰۱۷$	۰.۷۶۲۰
روش پیشنهادی	۰.۸۰۹۸

۵.۴ نتیجه گیری

درنهایت این سیستم می‌تواند منجر به ساخت یک شبکه‌ی عمیقی شود که می‌تواند عبارات را بر اساس معانی آن‌ها مدل کند. در این صورت می‌توانیم با سرچ در سند حاوی پاسخ و مقایسه با استفاده از اندازگیری میزان فاصله‌ی دو بردار پرسش با عبارات درون متن به پاسخ محتمل‌تر دست پیدا کنیم.

۶.۴ پیشنهادهای برخاسته از تحقیق

با توجه به تحقیقات حاصل از این پایان‌نامه می‌توان این نتیجه را گرفت که می توان با استفاده از کشف روابط میان کلمات درون عبارات کوتاه (پرسش‌ها و پاسخ‌ها) و مشخص کردن موضوع برای هر عبارت می‌توان برای هر عبارت قالبی را تعیین کرد تا بتوانیم جفت پرسش‌ها و پاسخ‌ها را شناسایی کنیم. این عملیات (تعیین موضوع با استفاده از کلمات کلیدی) که برروی بردارهای جاسازی عبارات و با نزیک کردن فاصله‌ی بردار عبارات با کلمات کلیدی صورت می‌گیرد، می‌تواند با اضافه کردن ویژگی‌های دستوری و معنایی دیگری بهبود یابد. همانطور که در پایان‌نامه اشاره شد، سیستم‌های جاسازی متنوعی توسط دانشمندان ساخته شده است که از جمله‌ی آن می‌توان به سیستم‌های جاسازی ساخته شده توسط دانشگاه استنفورد با نام $GloVe$ اشاره کرد که این سیستم جاسازی، متفاوت از $Word2vec$ عبارات را بر اساس ویژگی‌های دستوری‌شان مدل‌سازی می‌کند($Word2vec$ براساس ویژگی‌های معنایی کلمات را مدل‌سازی می‌کند). لذا از با استفاده از این ابزار می‌توان به جای آن که عبارات را بر اساس تنها ویژگی‌های معنایی و یا تنها دستوری مدل کنیم می‌توان از برادارهای دو بعدی

برای نمایش عبارات استفاده کنیم و هردو جاسازی را در کنار یکدیگر به کار ببریم تا بتوانیم آن چه را که پرسشگر از پرسش مطرح شده انتظار دارد را بهتر بتوانیم شناسایی کنیم تا بتوانیم بهتر به آن پاسخ دهیم.

ویژگی‌های دیگری نیز می‌تواند استخراج شود و در بازیابی اطلاعات درون سوال و پرسش استفاده شود تا بتوان پاسخ‌های صحیح، صریح و دقیق ارائه دهیم.

همچنین معیارهای فاصله که برای سنجش میزان فاصله‌ی دو بردار به کار گرفته می‌شود گاهی بسیار ناکارآمد می‌شود و بایستی از معیارهایی استفاده شود تا بتوانیم مقایسه‌های معنادار تری را برای مقایسه دو بردار عددی که حاوی اطلاعات و ویژگی‌ها متنی هستند ارائه دهیم.

۷.۴ جمع‌بندی کلی

در این پایان پایان‌نامه سعی داشتیم سیستم پرسش‌ـپاسخ اتوماتیکی ارئه دهیم تا بتوانیم بر اساس اطلاعیه‌هایی که به صورت سند در دسترس داریم بتوانیم به صورت خودکار به پرسش‌های مطرح شده پاسخ دهیم.

در سال ۲۰۱۷ جناب آقایی دکتر مینایی از دانشگاه نیویورک، با همکاری جناب آقای لی‌یو از شرکت $AT\&T$ مقاله‌ای [۱] را ارائه داده‌اند که در آن سعی شد سیستم پرسش‌ـ پاسخی ارائه شود تا با استفاده از شبکه‌های عصبی عمیق و اطلاعات داده شده در قالب سندهای از پیش تعریف شده، بتواند به پرسش‌های کاربران پاسخ داده شود.

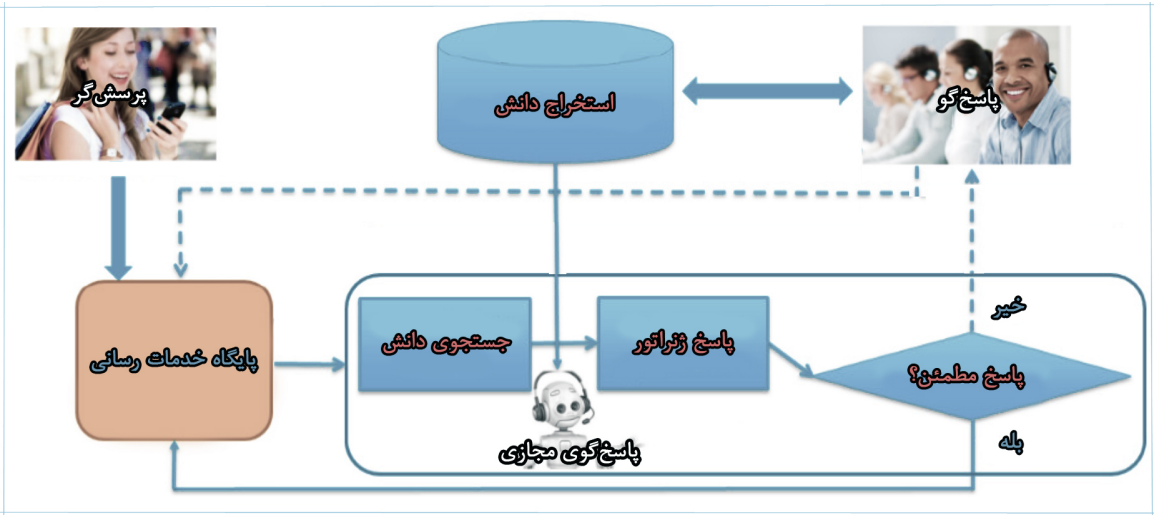
در این سیستم از شبکه‌های عصبی عمیق برای جاسازی پرسش‌ها استفاده می‌شود و در ادامه از بردارهایی که از این روش جاسازی بدست می‌آید برای آموزش شبکه‌ی عصبی استفاده می‌شود.

شبکه عصبی که در این مقاله استفاده شده است، شبکه‌ی عصبی سیام خوانده می‌شود، این شبکه که یک شبکه‌ی عصبی چندلایه است و براساس امتیازدهی به کاندیدهای پاسخ برای سوالات مطرح شده، بصورت نظارت شده و روش به روزرسانی وزن‌ها با استفاده از انتشار خطا به عقب، آموزش می‌بیند تا بتواند پاسخ مناسبی را برای پرسش های مطرح شده ارائه دهد.

در شبکه‌ی عصبی سیام، دو شبکه‌ی عصبی به صورت کاملا موازی ورودی دریافت می‌کنند، یکی پرسش‌ها را ورودی می‌کند و دیگری پاسخ‌ها را ورودی می‌کند. در ادامه پرسش‌ها و پاسخ‌ها در لایه‌های میانی و مخفی شبکه، استخراج ویژگی می‌شوند تا بتوان فاکتورهای سنجش شباهت برای هر جفت شناسایی و استخراج شود.

در ادامه ویژگی‌های استخراج شده از هر دو عبارت(جفت پرسش‌ـ پاسخ) با یکدیگر مقایسه می‌شوند و پرسش‌ـ پاسخ‌های متناظر مشخص می‌شوند. این روند تا آنجا ادامه پیدا می‌کند تا وزن‌های شبکه‌ی عصبی به درستی ساخته شوند در انتها از این وزن‌ها استفاده می‌شود تا برای پرسش‌های مطرح شده، پاسخ معتبری از درون سند پیدا شود.

این سیستم می‌تواند به عنوان یک همیار اپراتور در پاسخگویی به پرسشگران یک مرکز بکار



شکل ۶.۴: **نمایی از یک سیستم پاسخگویی**

گرفته شود به این صورت که ابتدا پرسش مطرح شده از جانب کاربر تحلیل می شود و سپس در درون سند به جستجوی پاسخ پرداخته می‌شود، در صورتی که در سند پاسخی با امتیاز شباهت کافی (بصورت دستی و بر اساس نوع کاربرد تنظیم می‌شود) وجود داشته باشد، پاسخ را به کاربر منتقل می‌کند و در غیر این صورت پرسش را به اپراتور برای بررسی انتقال می‌دهد شکل ۶.۴.

الگوریتم خلاقانه‌ی جناب آقای مینایی، برای جاسازی عبارات (پرسش‌ها و پاسخ‌ها) از روش $doc2vec$ استفاده کرده است. این روش جاسازی به درستی بردارهایی از عبارات را ارائه می‌دهد که این بردارها از نظر فاصله در فضای برداری، بردارهای معناداری ایجاد می‌کند. در این پایان‌نامه با ارائه‌ی روشی برای جاسازی درست عبارات بر اساس تاثیر کلمات کلیدی در هر عبارت سعی شد تا عباراتی را که از لحاظ معنی و موضوع بایکدیگر نزدیکی دارند، بردارهای نزدیک به همی در فضای برداری داشته باشند.

در این روش ابتدا بردار $doc2vec$ هر عبارت و بردار $word2vec$ کلمات درون هر عبارت به دست آورده می‌شود. سپس براساس میزان اهمیت هر کلمه، ضریبی برای هر کلمه مشخص می‌شود و در نهایت بردار عبارت براساس میانگین وزنی بردار عبارت $doc2vec$ و بردارهای $word2vec$ کلمات ساخته می‌شود.

این جاسازی عبارات، موجب می‌شود تا جفت پرسش‌ـپاسخ‌ها به یکدیگر شبیه شوند و اینگونه شبکه‌ی عصبی سیام می‌تواند سریع‌تر و دقیق‌تر جفت‌ها متناظر را شناسایی کند.

ما با استفاده از این روش توانستیم خطای حاصل از کار آقای مینایی را تا میزان ۲۰ درصد کاهش دهیم.

مراجع

[1] S.Minaee, Z.Liuy, ”Automatic Question-Answering Using A Deep Similarity Neural Network”, IEEE, 2017.

[2] BF Green Jr, AK Wolf, C Chomsky, K Laughery, “Baseball: an automatic question-answerer”, western joint IREAIEE-ACM computer conference, ACM, 1961.

[3] Robert F. Simmons ,”Natural language question-answering systems”, Communications of the ACM 13(1):15-30, 1969.

[4] Robert Wilensky, David N. Chin Marc Luria, James Martin, James Mayfield, and Dekai Wu, ”The Berkeley UNIX Consultant Project” Computational Linguistics, Volume 14, Number 4, 1988.

[5] Collobert, R., Weston, J, ”A unified architecture for natural language processing: Deep neural networks with multitask learning”, In Proceedings Of The 25th International Conference On Machine Learning, pp, 160–167, ACM, 2008.

[6] Chen, D., & Manning, C, ”A Fast and Accurate Dependency Parser using Neural Networks”, In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 740-750, Doha, Qatar. Association for Computational Linguistics, 2014.

[7] Weiss, D., Alberti, C., Collins, M., & Petrov, S, ”Structured Training for Neural Network Transition-Based Parsing”, In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 323-333, Beijing, China. Association for Computational Linguistics, 2015.

[8] Pei, W., Ge, T., & Chang, B, ”An Effective Neural Network Model for Graph-based Dependency Parsing”, In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing

(Volume 1: Long Papers), pp. 313-322, Beijing, China. Association for Computational Linguistics, 2015.

[9] Durrett, G., & Klein, D, ”Neural CRF Parsing”, In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 302-312, Beijing, China. Association for Computational Linguistics, 2015.

[10] Lewis, M., & Steedman, M, ”Improved CCG Parsing with Semi-supervised Supertagging”, Transactions of the Association for Computational Linguistics, 2(0), 327-338, 2014.

[11] Henderson, M., Thomson, B., & Young, S, ”Deep Neural Network Approach for the Dialog State Tracking Challenge”, In Proceedings of the SIGDIAL 2013 Conference, pp. 467-471, Metz, France. Association for Computational Linguistics, 2013.

[12] de Gispert, A., Iglesias, G., & Byrne, B, ”Fast and Accurate Preordering for SMT using Neural Networks”, In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1012-1017, Denver, Colorado. Association for Computational Linguistics, 2015.

[13] Bengio, Y., Ducharme, R., Vincent, P., & Janvin, C, ”A Neural Probabilistic Language Model”, J. Mach. Learn. Res., 3, 1137-1155, 2003.

[14] Vaswani, A., Zhao, Y., Fossum, V., & Chiang, D, ”Decoding with Large-Scale Neural Language Models Improves Translation” In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp. 1387-1392, Seattle, Washington, USA. Association for Computational Linguistics, 2013.

[15] Iyyer, M., Boyd-Graber, J., Claudino, L., Socher, R., & Daum´e III, H, ”A Neural Network for Factoid Question Answering over Paragraphs”, In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 633-644, Doha, Qatar. Association for Computational Linguistics, 2014.

[16] Johnson, R., & Zhang, T, ”Effective Use of Word Order for Text Categorization with Convolutional Neural Networks”, In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 103-112, Denver, Colorado. Association for Computational Linguistics, 2015.

[17] Elman, J. L, ”Finding Structure in Time”, Cognitive Science, 14(2), 179-211, 1990.

[18] Goller, C., & K¨uchler, A, ”Learning Task-Dependent Distributed Representations by Back-propagation Through Structure”, In In Proc. of the ICNN-96, pp. 347-352. IEEE, 1996.

[19] Mikolov, T., Karafi´at, M., Burget, L., Cernocky, J., & Khudanpur, S, ”Recurrent neural network based language model”, In INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September 26-30, 2010, pp. 1045–1048, 2010.

[20] Mikolov, T., Kombrink, S., Luk´aˇs Burget, ˇCernocky, J. H., & Khudanpur, S, ”Extensions of recurrent neural network language model. In Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, pp. 5528–5531. IEEE, 2011.

[21] Mikolov, T, ”Statistical language models based on neural networks”, Ph.D. thesis, Ph. D. thesis, Brno University of Technology, 2012.

[22] Duh, K., Neubig, G., Sudoh, K., & Tsukada, H, ”Adaptation Data Selection using Neural Language Models: Experiments in Machine Translation”, In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 678–683, Sofia, Bulgaria. Association for Computational Linguistics, 2013.

[23] Adel, H., Vu, N. T., & Schultz, T, ”Combination of Recurrent Neural Networks and Factored Language Models for Code-Switching Language Modeling”, In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 206–211, Sofia, Bulgaria. Association for Computational Linguistics, 2013.

[24] Auli, M., Galley, M., Quirk, C., & Zweig, G, ”Joint Language and Translation Modeling with Recurrent Neural Networks”, In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp. 1044–1054, Seattle, Washington, USA. Association for Computational Linguistics, 2013.

[25] Auli, M., & Gao, J, ”Decoder Integration and Expected BLEU Training for Recurrent Neural Network Language Models” In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 136–142, Baltimore, Maryland. Association for Computational Linguistics, 2014.

[26] Xu, W., Auli, M., & Clark, S, ”CCG Supertagging with a Recurrent Neural Network”, In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), pp. 250–255, Beijing, China. Association for Computational Linguistics, 2015.

[27] Ling, W., Dyer, C., Black, A. W., Trancoso, I., Fermandez, R., Amir, S., Marujo, L., & Luis, T, ”Finding Function in Form: Compositional Character Models for Open Vocabulary Word Representation”, In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 1520–1530, Lisbon, Portugal. Association for Computational Linguistics, 2015.

[28] Iyyer, M., Enns, P., Boyd-Graber, J., & Resnik, P, ”Political Ideology Detection Using Recursive Neural Networks”, In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1113–1122, Baltimore, Maryland. Association for Computational Linguistics, 2014.

[29] Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P, ”Natural language processing (almost) from scratch”, The Journal of Machine Learning Research, 12, 2493–2537, 2011.

[30] Chen, D., & Manning, C, ”A Fast and Accurate Dependency Parser using Neural Networks”, In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 740–750, Doha, Qatar. Association for Computational Linguistics, 2014.

[31] Mitchell Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini, ”Building a large annotated corpus of English : The PennTreebank”, Computational Linguistics 19,2(1993),313–330, 1993.

[32] Ahmed Magdy Ezzeldin and Mohamed Shaheen, ”A survey of Arabic question answering: challenges, tasks, approaches, tools, and future trends”, In International Arab Conf on Information Technology. 1–8, 2012.

[33] Cordell Green, ”Theorem proving by resolution as a basis for question-answering systems”, Machine Intelligence 4 (1969), 183–205, 1969.

[34] Wenhui Wang, Nan Yang, Furu Wei, Baobao Chang, and Ming Zhou, ”Gated self-matching networks for reading comprehension and question answering”, In Annual Meeting of the ACL, Vol. 1. 189–198, 2017.

[35] Li Dong, Furu Wei, Ming Zhou, and Ke Xu, ”Question answering over freebase with multi-column convolutional neural networks”, In Annual Meeting of the ACL and International Joint Conf on Natural Language Processing, Vol. 1. 260–269, 2015.

[36] David Raposo, Adam Santoro, David Barrett, Razvan Pascanu, Timothy Lillicrap, and Peter Battaglia, ”Discovering objects and their relations from entangled scene representations”, arXiv preprint arXiv:1702.05068, 2017.

[37] Adam Santoro, David Raposo, David G Barrett, Mateusz Malinowski, Razvan Pascanu, Peter Battaglia, and Tim Lillicrap, ”A simple neural network module for relational reasoning”, In Advances in neural information processing systems. 4974–4983, 2017.

[38] Chopra, S.; Hadsell, R.; LeCun, Y, ”Learning a similarity metric discriminatively, with application to face verification”. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05). 1: 539–546 vol. 1, 2005.

[39] Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L, ”DeepFace: Closing the Gap to Human-Level Performance in Face Verification”. 2014 IEEE Conference on Computer Vision and Pattern Recognition: 1701–1708, 2014.

[40] Williams, Ronald J.; Hinton, Geoffrey E.; Rumelhart, David E, ”Learning representations by back-propagating errors”. Nature. 323 (6088): 533–536. doi:10.1038/323533a0. ISSN 1476-4687, 1986.

[41] Klaus Greff, Rupesh K Srivastava, Jan Koutník, Bas R Steunebrink, and Jürgen Schmidhuber, ”LSTM: A search space odyssey”, IEEE Transactions on Neural Networks and Learning Systems 28, 10 (2017), 2222–2232, 2017.

[42] Sepp Hochreiter and Jürgen Schmidhuber, ”Long short-term memory”, Neural computation 9, 8 (1997), 1735–1780, 1997.

[43] T Mikolov, I Sutskever, K Chen, GS Corrado, J Dean, ”Distributed representations of words and phrases and their compositionality”, In Advances in neural information processing systems, 2013.

[44] G Hinton, JL McClelland, DE Rumelhart, ”Distributed representations”, In Parallel distributed processing: Explorations in the microstructure of cognition, Foundations, MIT Press, 1986.

[45] Y Bengio, R Ducharme, P Vincent, C Jauvin, ”A neural probabilistic language model”, journal of machine learning research, pp.1137-1155, 2003.

[46] CD Manning, H Schutze, ”Foundations of statistical natural language processing”, Vol. 999. Cambridge: MIT press, 1999.

[47] T Mikolov, I Sutskever, K Chen, GS Corrado, J Dean, ”Distributed representations of words and phrases and their compositionality”, In Advances in neural information processing systems, 2013.

[48] T Mikolov, K Chen, G Corrado, J Dean, ”Efficient estimation of word representations in vector space”, arXiv preprint arXiv:1301.3781, 2013.

[49] S Chopra, R Hadsell, Y LeCun, ”Learning a similarity metric discriminatively, with application to face verification”, Conference on Computer Vision and Pattern Recognition, IEEE, 2005.

[50] J Bromley, JW Bentz, L Bottou, I Guyon, Y LeCun, C Moore, E Sackinger, R Shah, ”Signature verification using a Siamese time delay neural network”, International Journal of Pattern Recognition and Artificial Intelligence, 1993.

[51] J Bromley, JW Bentz, L Bottou, I Guyon, Y LeCun, C Moore, E Sackinger, R Shah, ”Signature verification using a Siamese time delay neural network”, International Journal of Pattern Recognition and Artificial Intelligence, 1993.

[52] James Allen, ”Natural Language Understanding”, Benjamin/Cummings Publishing Company, 1995 .

[53] Terry Winograd, ”Language as a Cognitive Process”, Addison-Wesley, 1983.

[54] Eugene Charniak, ”Statistical language learning”, MITpress, 1996.

[55] Dan Jurafskyand James Martin, ”Speech and language processing”, Pearson Education, 2000.

[56] Anthony Valiant Phillips, ”A Question-answering Routine”, MITpress, 1960.

Aabstract

Over the last few years, researchers in the field of natural language processing have been making great progress through the use of deep learning models, and Q & A systems as a subset of the science of natural language processing have not been exempted from this. In this thesis, we intend to provide a deep learning-based model for automatic answer to the question. With this in mind, we can divide the question-answer system into three parts: categorizing questions and answers, retrieval information, and choosing an answer, first we classify the questions and answers using a neural-probabilistic modeling and transform into a vector, then We teach a deep neural network similarity to find out the similarity between questions and answers, and finally we select the best answer. Our proposed system, by proposing a method based on Neural probabilistic modeling, transforms the expression into a vector, so that the vector of each statement become very similar to the word vector or its keywords so that the likelihood of the similarity of the question with its corresponding answers could raise and finally siamese neural network, with changes, learns in a way that it can choose the right answer for the question. In this model, we managed to reduce the error by about 25% compared to the current one.

keywords : artificial intelligence, natrual language processing, question-answering system, deep learning tools, artificial neural network, word2vec, deep similarity neural network.

Shahrood University Of Technology

Computer Engineering And Information Technology

MSc Thesis in: Artificial Intelligence

Question-Answering System And Deep Learning Tools

By: Navid Soleymani

Supervisor

Morteza Zahedi

Advisor

Hamid Hasanpor

June 2019