تاریخ دریافت مقاله: ۹۸/۰۳/۲۶
تاریخ پذیرش مقاله: ۹۸/۰۹/۲۰

تشخیص موجودیت‌های اسمی بر پایه شبکه‌های عصبی با حافظه بلند کوتاه مدت پایدار

نعیمه علی پور
دانشجوی کارشناسی ارشد، دانشکده مهندسی فناوری اطلاعات و کامپیوتر ـ دانشگاه صنعتی ارومیه ـ ارومیه ـ ایران
پست الکترونیکی: naimeh.alipour@it.uut.ac.ir

جعفر طهمورث نژاد*
استادیار دانشکده مهندسی فناوری اطلاعات و کامپیوتر ـ دانشگاه صنعتی ارومیه ـ ارومیه ـ ایران
پست الکترونیکی: j.tahmores@it.uut.ac.ir

چکیده:

مسئلهٔ شناسایی موجودیت‌های اسمی به‌عنوان یکی از شاخه‌های پردازش زبان طبیعی و زیرمجموعه‌ای از استخراج اطلاعات به‌شمار می‌رود. هدف اصلی در تشخیص موجودیت‌های اسمی، دسته‌بندی اسامی خاص متن با برچسب‌هایی مانند شخص، مکان و سازمان است. امروزه اکثر روش‌هایی که برای این منظور معرفی شده‌اند سعی در استفاده از ویژگی‌هایی دارند که مختص به یک زبان خاص نیست. از جمله این روش‌ها می‌توان به استفاده کوتاه مدّت پایدار اشاره‌کرد. آموزش این شبکه‌های عصبی به‌وسیله استخراج ویژگی‌ها از بردارهای کلمات در سطح نویسه و بردارهای کلمات از پیش آموزش دیده انجام می‌گیرد. دراین مقاله مدل جدیدی برای ساخت نمایش‌های برداری کلمات پیشنهاد می‌شود که از ترکیب بردار نحوی کلمه با دو بردار کلمه قبلی به‌دست می‌آید. بردار نحوی کلمه شامل اطلاعات نحوی موجود در جمله مانند موقعیت کلمه در جمله، نقش کلمات و ارتباط نحوی آن‌ها با یکدیگر است. استفاده از این روش باعث توسعه سیستم‌هایی می‌شود که کمترین وابستگی را به دامنه دارند. کارایی روش پیشنهادی بر روی مجموعه داده موجودیت‌های اسمی مورد ارزیابی قرار گرفته‌است. نتایج به‌دست آمده، نشان‌دهنده بهبود قابل ملاحظه‌ای در مقایسه با سایر روش‌های مطرح در حوزه تشخیص موجودیت‌های اسمی می‌باشد.

واژه‌های کلیدی: پردازش زبان طبیعی، تشخیص موجودیت‌های اسمی، نمایش برداری کلمات، شبکه عصبی با حافظه بلند مدت – کوتاه مدّت پایدار

۱- مقدمه

شناسایی موجودیت‌های اسمی[1] به‌عنوان حوزه‌ای از پردازش زبان طبیعی و زیرمجموعه‌ای از استخراج اطلاعات به‌شمار می‌رود. در فرآیند تشخیص موجودیت‌های اسمی هدف این است که کلیه اسامی خاص موجود در متن، شناسایی و در دسته‌های از پیش تعیین شده‌ای مانند اسامی اشخاص، سازمان‌ها و مکان‌ها طبقه‌بندی گردند [۱].

* نویسندهٔ مسئول

1- Name Entity Recognition (NER)

سیستم‌های تشخیص موجودیت‌های اسمی در سیستم‌های استخراج اطلاعات[2]، سیستم‌های پرسش و پاسخ[3]، طبقه‌بندی متون[4]، خلاصه‌سازی متون و بهینه‌سازی جستجو کاربرد دارند [۲] [۳].

سیستم‌های تشـــخیص موجودیت‌های اسمی همواره با چالش‌هایی مواجه بودند، زیرا در بســـیاری از زبان‌ها، مقدار کافی داده طبقه‌بندی شده برای آموزش آن‌ها موجود نیســـت. از ســـویی دیگر، در اکثر زبان‌ها محدودیت‌هایی در نوع کلماتی که می‌توانند به‌عنوان اســـم شناخته شوند وجـــود دارد. به همین دلیل از ویژگی‌های زبان شـــناختی مختص بـــه یک زبان مانند ویژگی‌های کلمه‌ای، فهرســـتی و ســـندی برای آموزش این سیســـتم‌ها استفاده می‌شود [٤]. روش‌هایـــی کـــه از این ویژگی‌های اســـتفاده می‌کنند، روش‌های وابسته به زبان نامیده می‌شوند [٥]. به‌طور کلی این روش‌ها به ســـه دســـته زیر تقسیم‌بندی می‌شوند [٦]: ۱) روش‌های مبتنی بر قانون: در این روش سعی می‌شود قالب و شـــکل کلی موجودیت‌های اسمی را به‌صورت یک عبارت با قاعده نمایش داد تا ســـامانه، اسامی خاص را بر مبنای این عبارت‌ها تشـــخیص دهد. در واقع در این روش موجودیت‌های اســـمی به وســـیله مولفه‌هایی که در ظاهر عبارت‌ها ممکن اســـت وجود داشته‌باشـــد تشخیص داده می‌شـــوند. برای مثال در زبان انگلیسی عبارت‌هایی که با کلمات یا حروفی از قبیل Dr یا Mr شـــروع می‌شوند، اسم یک شخص خواهندبود. ۲) روش‌های مبتنی بر واژه‌نامه: در این روش یک لیســـت از موجودیت‌های اسمی و نوع آن‌ها تهیه شـــده و موجودیت‌های اســـمی داخل متن با استفاده از این لیســـت تشخیص داده می‌شـــود. برای مثال اسامی مکان‌ها در فرهنگ‌های جغرافیایی که شامل میلیون‌ها اسم مکان به همراه مشخصات جغرافیایی آن‌ها هستند، به‌عنوان یک واژه‌نامه مناســـب برای تشخیص اسامی مکان‌ها قابل استفاده است. ۳) روش‌های آماری: این روش اغلب به‌عنوان روش اصلی در تشخیص اسامی خاص شناخته می‌شود و ســـایر روش‌ها به‌عنوان مکمل در کنار آن مورد اســـتفاده قرار می‌گیرند. در این روش ابتدا سیستم به وسیله پیکره‌ای از داده‌های آموزشی که به‌صورت دستی و به‌وسیله انسان برچسب‌گذاری شده‌اند، آموزش دیده و سپس با یادگیری از این داده‌ها به تشـــخیص خودکار اسامی خاص در متن می‌پردازد.

با این حال ســـاخت منابع بـــا ویژگی‌های مختص یک زبان هزینه زیـــادی در بردارد. به همیـــن دلیل روش‌های جدیدی تحت عنوان روش‌های مستقل از زبان برای ساخت سیســـتم‌های تشـــخیص موجودیت‌های اســـمی پیشنهاد شده‌اســـت [۷]. از جمله این روش‌ها استفاده از شبکه‌های عصبی است که به هیچ یک از ویژگی‌ها و منابع مختص یک زبان نیازی ندارد [۸]. شـــبکه‌های عصبی معرفی شده در این حوزه می‌توانند تنها با استفاده از داده‌های برچسب‌دار و طبقه‌بندی شـــده به یادگیری موجودیت‌های اســـمی یک زبـــان جدید بپردازند و پس از آن، کار طبقه‌بندی اســـامی را به راحتی انجام‌دهند [۹]. در این مقاله از شـــبکه عصبی کوتاه‌مدت – کوتاه مدّت پایدار[5] با میدان شرطی تصادفی[6] LSTM-CRF برای تشخیص موجودیت‌های اسمی استفاده شده‌است.

استفاده از شـــبکه‌های عصبی چالش دیگری را مطرح می‌کنـــد. در واقـــع از آنجایـــی کـــه شـــبکه‌های عصبی از ویژگی‌های وابســـته به زبان اســـتفاده نمی‌کنند، بنابراین باید به دنبال یک ســـری ویژگی‌های دیگری برای آموزش و آزمایش آن‌ها بود. بدیـــن منظور نمایش برداری کلمات به‌عنوان ورودی شـــبکه‌های عصبی معرفی می‌شـــوند که به اســـتخراج ویژگی‌هایی در سطح نویســـه برای هر یک از کلمـــات موجود در جمله می‌پردازنـــد [۱۰]. در این مقاله سعی شده‌است علاوه بر ویژگی‌های ذکر شده، از اطلاعات نحوی کلمه موجود در جمله نیز در ساخت نمایش برداری کلمات استفاده شود. استفاده از این اطلاعات در تشخیص بهتر موجودیت‌های اسمی تاثیر بسیاری دارد. این اطلاعات

2- Information Extraction
3- Question Answering
4- Text Categorization
5- Long short-term memory networks (LSTM)
6- Conditional Random Field (CRF)

نحوی می‌تواند شـــامل نقش کلمات (اســـم، فعل و صفت)، موقعیت کلمه در جمله و یا ارتباط نحوی کلمات با یکدیگر باشد.

همچنیـــن در این مقاله برای بهبود عملکرد شـــبکه‌های عصبی بـــا حافظه کوتاه مدّت پایـــدار، مدل نمایش برداری جامع کلمات پیشـــنهاد می‌شـــود. در این مدل ابتدا برای هر کلمه یک نمایش برداری با استفاده از نویسه‌های آن ساخته می‌شـــود. ســـپس بردار معادل آن کلمه از جدول کلمات از پیش آموزش دیده، جســـتجو می‌شود. ســـپس برای کلمه مورد نظر یک بـــردار نحوی با توجه به اطلاعات نحوی آن در جمله ســـاخته می‌شود. در نهایت این ســـه بردار با هم ترکیب شـــده و بردار جامع کلمات را تشکیل می‌دهند که به عنوان ورودی به شـــبکه عصبی داده می‌شـــود تا برچسب اسمی کلمه مشـــخص شود. مدل پیشـــنهادی بر روی یک پیکـــره متنی انگلیســـی آزمایش و اجرا شده‌اســـت ولی به دلیل مســـتقل از زبان بودن روش، قابلیت استفاده در سایر زبان‌هـــا را نیز دارد. تشـــخیص موجودیت‌های اســـمی با استفاده از شـــبکه‌های عصبی با حافظه کوتاه مدّت پایدار در این مقاله، عملکرد بهتری را نســـبت به سایر روش‌های وابسته به زبان نشان می‌دهد. همچنین کارایی این شبکه‌های عصبی با اســـتفاده از نمایش برداری مدل پیشنهادی بهبود می‌یابد. مقایسه نتایج به‌دســـت آمده از مدل ما با مدل‌های دیگر بیانگر این اســـت که روش ما می‌تواند موجودیت‌های اســـمی جدید و دارای ساختار غیر معمول را نیز شناسایی کند.

ادامه مقاله به‌صورت زیر ســـازماندهی شده‌است. در بخش ۲ پژوهش‌های مرتبط مورد بررسی قرار می‌گیرند. در بخـــش ۳ شـــبکه عصبی اصلـــی مورد اســـتفاده برای تشخیص موجودیت‌های اسمی معرفی می‌شود. در بخش ٤ مدل پیشنهادی شرح داده می‌شود. در بخش ٥ آزمایش‌های انجام گرفته ارائه شده‌است. در بخش ٦ به ارزیابی مدل و نتایج پرداخته می‌شود. در پایان نتیجه‌گیری و کارهای آتی آورده شده‌است.

۲- کارهای پیشین

به‌طور کلی روش‌های ارائه شـــده در حوزه تشخیص موجودیت‌های اسمی را می‌توان به سه دسته اصلی تقسیم بنـــدی کرد: ۱- روش‌های دســـته اول با عنـــوان مدل‌های وابسته به زبان با استفاده از ویژگی‌های زبان شناختی و فرهنگ‌های جغرافیایی به تشـــخیص موجودیت‌های اسمی می‌پردازنـــد. از جمله کارهـــای انجام گرفتـــه در این مدل می‌توان به کار لین و وو اشـــاره‌کرد. آن‌هـــا از یک میدان شرطی تصادفی زنجیره‌ای خطی استفاده کردند که به آن ویژگی‌های خوشه‌های عبارات اضافه شدند. این ویژگی‌ها از داده‌های وب و املای کلمات اســـتخراج می‌شـــوند [۱۱]. پاسوس و همکاران نیز از یک میدان شرطی تصادفی خطی به همراه ویژگی‌های املایی و فرهنگ جغرافیایی گازیتیرز[۷] اســـتفاده نمودند [۱۲]. راتینو و روث نیز به مقایسه چندین رویکرد برای تشخیص موجودیت‌های اسمی پرداخته‌اند و مدل نظارت شده خود را با استفاده از پرسپترون میانگین تنظیم شده و جمع آوری اطلاعات زمینه‌ای می‌سازند [۱۳]. ۲- روش‌هـــای دســـته دوم با عنوان مدل‌های شـــبکه‌های عصبی هســـتند که از شـــبکه‌های عصبی مختلف معرفی شده در حوزه پردازش زبان طبیعی استفاده می‌کنند. برای مثال کولوبرت و همکاران از یک شبکه عصبی کانوولوشن بـــر روی دنباله‌ای از بردارهای کلمـــات، به همراه یک لایه CRF بر روی آن اســـتفاده کردنـــد [۱٤]. هانگ و همکاران مدل شـــبکه‌های عصبی LSTM به همراه یـــک لایه CRF را ارائه دادند، که از ویژگی‌های املایی کلمات استفاده می‌کند [۱٥]. ژو و همکاران نیز از مدلی مشـــابه استفاده کردند و آن را با وظیفه برچســـب‌زنی نقـــش معنایی تطبیق داده‌اند [۱٦]. ۳- روش‌های دسته سوم مدل‌های شبکه‌های عصبی مبتنی بر نویسه نام دارند، که در آن‌ها از بردارهای کلمات به‌عنوان ورودی‌های شـــبکه‌های عصبی مختلف استفاده می‌شود. استفاده از نمایش‌های برداری کلمات برای اولین بار توســـط کولبرت و همکاران مطرح شـــد. آن‌ها مدلی از شـــبکه‌های عصبی پیش‌خور را پیشنهاد دادند که به یافتن

7- gazetteers

ویژگی‌ها از بردارهای کلماتی می‌پردازند که بر روی مقدار بزرگی از داده‌ها آموزش دیده‌اند [۱٤]. در حال حاضر اکثر مدل‌هایی که برای تشـــخیص موجودیت‌های اسمی معرفی می‌شـــوند از نمایش‌های برداری مبتنی بر حروف استفاده می‌کننـــد. در این نوع مدل‌ها هدف اســـتخراج ویژگی‌ها از بردارهای کلمات در ســـطح نویســـه اســـت. چیو و نیکولز به معرفی شـــبکه‌های عصبی کانوولوشن برای استخراج ویژگی‌های لازم برای آموزش شبکه عصبی LSTM مورد استفاده برای تشـــخیص موجودیت‌های اسمی، پرداختند [۱۷] . هووی و همکاران [۱۸] و ســـنتوس و همکاران [۱۹] نیز رویکردی مشـــابه را به‌کار برده‌اند. لمپل و همکاران از شـــبکه‌های عصبی Bi-LSTM برای استخراج ویژگی‌ها و از شـــبکه LSTM-CRF برای تشـــخیص موجودیت‌های اسمی اســـتفاده نمودنـــد [۲۰]. لینگ و همـــکاران [۲۱] نیز روش مشابهی را برای برچسب‌گذاری اجزای سخن به‌کار بردند.

روش پیشنهادی در این مقاله، بر پایه شبکه‌های عصبی مبتنی بر نویســـه بوده که در آن از شـــبکه‌های عصبی با حافظه کوتاه مدّت پایدار اســـتفاده می‌شـــود. سپس برای بهبود عملکرد این شـــبکه‌های عصبی، مدل نمایش برداری جامع کلمات پیشـــنهاد می‌شـــود. در ســـاخت این نمایش برداری علاوه بر اســـتفاده از بردارهای کلمات در ســـطح نویســـه و بردارهای از پیش آموزش دیده، از بردار نحوی کلمه نیز استفاده شده است.

۳- مدل LSTM-CRF

شـــبکه‌های عصبی بازگشـــتی(RNN)[8] خانـــواده‌ای از شبکه‌های عصبی هستند که برای داده‌های دنباله‌دار مورد استفاده قرار می‌گیرند. آن‌ها دنباله‌ای از داده‌های ورودی را می‌گیرند و پس از پردازش دنباله دیگری را که در بردارنده اطلاعاتی در مورد رشـــته مورد نظر اســـت برمی‌گردانند. تئوری شبکه‌های RNN بر این اساس است که این شبکه‌ها می‌توانند وابستگی‌های طولانی بین دنباله‌ها را یاد بگیرند، ولی در عمل این امر ممکن نیست و با وارد شدن دنباله‌های جدید، دنباله‌های قدیمی‌تر از بین می‌روند [۲۲].

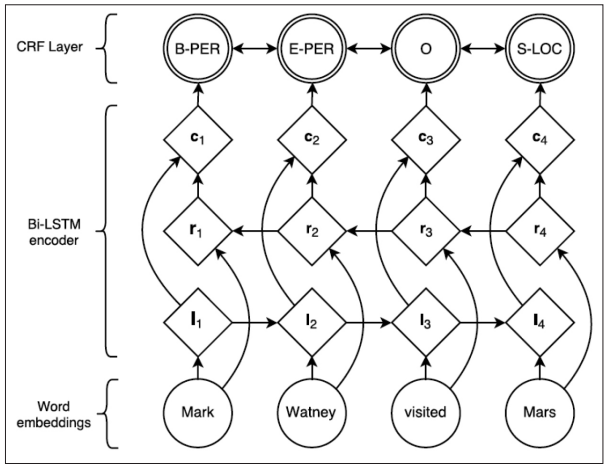
برای حل این مشکل شبکه‌های عصبی با حافظهٔ کوتاه مدّت پایدار که از این پس به اختصار در این مقاله به آن‌ها LSTM می‌گوییم، طراحی و پیشـــنهاد شدند. این شبکه‌ها با اســـتفاده از چندین ســـلول حافظه امکان به خاطر سپاری دنباله‌های بیشتری از ورودی‌ها و کشف روابط بین آن‌ها را فراهم می‌کنند [۲۳]. امروزه نیز استفاده از شبکه‌های LSTM به دلیل ویژگی‌هایی که دارند در تشخیص موجودیت‌های اسمی استفاده گسترده‌ای پیدا کرده است [۲۰] .

برای یک جمله ورودی (x1,x2,...,xn) که n کلمه[9] دارد، LSTM برای هر کلمه که با t نمایش داده می‌شود، یک نمایش از زمینه[10] ســـمت چپ کلمه به‌عنوان $\overrightarrow{h_t}$ محاسبه می‌کند. همچنین یک نمایش از زمینه ســـمت راست کلمه با عنوان $\overleftarrow{h_t}$ نیز توســـط یک شـــبکه LSTM دیگر با خواندن کلمات موجود در جمله به‌صورت معکوس از آخر به اول، به‌دست می‌آید. به شبکه عصبی اول LSTM رو به جلو[11] و به شبکه عصبی دوم LSTM رو به عقب[12] گفته می‌شود. هردوی این LSTM ها شـــبکه‌هایی مجزا با پارامترهای مختلف هستند کـــه به آن‌هـــا LSTM دو جهتـــه[13] نیز گفته می‌شـــود [۲٤]. در نهایـــت خروجی این شـــبکه عصبـــی LSTM دو جهته، نمایشی از کلمه اســـت که از الحاق نمایش زمینه‌های چپ و راست آن به‌دســـت می‌آید، $h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}]$. این خروجی حاوی اطلاعات مفیدی از کلمه اســـت که می‌تواند توســـط برنامه‌های مختلف برچســـب‌گذاری مورد اســـتفاده قرار بگیرد.

بدین ترتیـــب، می‌توان از این نمایـــش کلمات به‌عنوان ویژگی‌هایی[14] برای انجام تصمیمات برچسب‌گذاری برای هر خروجی y_t به‌صورت مستقل استفاده‌کرد [۲۵]. اگرچه این مدل در مسائل ساده برچسب‌گذاری به خوبی عمل می‌کند ولی گاهی اوقات ممکن است ارتباط قوی بین برچسب‌های

8- Recurrent neural networks (RNNs)

9-Term
10-Context
11-Forward LSTM
12-Backward LSTM
13-Bidirectional LSTM
14-Feature



شکل ۱: معماری مدل LSTM-CRF برای تشخیص موجودیت‌های اسمی[۲۰]. کلمات جمله مورد نظر در قالب بردارهایی به شبکه عصبی LSTM رو به جلو و شبکه عصبی LSTM رو به عقب داده می‌شوند. l_i نشان‌دهنده کلمه iام و زمینه سمت چپ آن و r_i نیز نشان‌دهنده کلمه iام و زمینه سمت راست آن است.

در نهایت این خروجی‌ها در قالب بردار c_i با هم ترکیب شده و برای محاسبه امتیازها و مشخص کردن برچسب موجودیت اسمی به لایه CRF داده می‌شوند.

خروجی وجود داشته باشد، در نتیجه نمی‌توان به‌صورت دقیق ردهٔ آن کلمه را مشخص کرد. مسئله تشخیص موجودیت‌های اسمی یک نوع از این مسائل است که در آن یک‌سری محدودیت‌هایی بر روی برچسب‌ها وجود دارد و ترکیبات خاصی از برچسب‌های خروجی ممکن نیستند و هرگز نباید اتفاق بیفتد. به‌عنوان مثال اگر اسم یک مکان را در ابتدای جمله داشته باشیم، اسم یک شخص نمی‌تواند به دنبال آن بیاید. بنابراین از یک لایه شبکه عصبی دیگر به نام میدان شرطی تصادفی[۱۵] (CRF) در انتهای شبکه Bi-LSTM استفاده می‌شود که می‌تواند پیش‌بینی‌های مربوط به برچسب منتشر شده را بر مبنای کل دنباله خروجی انجام دهد و قواعدی مانند قاعده قبل را شناسایی کند و بر اساس آن تصمیمات برچسب‌گذاری را انجام دهد. در نتیجه به‌جای این‌که تصمیمات برچسب‌گذاری به‌صورت مستقل انجام شوند، به‌صورت مشترک صورت می‌گیرد [۲۶] . شکل شماره ۱ معماری مدل LSTM-CRF را برای جمله زیر نشان می‌دهد:

Mark Watney visited Mars.

که کلمات جمله مورد نظر در قالب بردارهایی به شبکه عصبی LSTM رو به جلو و شبکه عصبی LSTM رو به عقب داده می‌شوند. l_i نشان‌دهنده کلمه iام و زمینه سمت چپ آن و r_i نیز نشان‌دهنده کلمه iام و زمینه سمت راست آن است. در نهایت این خروجی‌ها در قالب بردار c_i با هم ترکیب شده و برای محاسبه امتیازها و مشخص کردن برچسب موجودیت اسمی به لایه CRF داده می‌شوند.

برای یک جمله ورودی $X=(x_1,x_2,...,x_n)$ ماتریسی به نام که امتیازهای خروجی شبکهٔ دو طرفه LSTM را نشان می‌دهد، در نظر می‌گیریم. اندازه این ماتریس برابر با n×k است، که تعداد برچسب‌ها و P_{ij} امتیاز برچسب jام به کلمه iام را در یک جمله نشان می‌دهد. در واقع برای دنباله‌ای از کلمات خروجی $Y=(y_1,y_2,...,y_n)$، امتیازهای پیش‌بینی شده برای هر کلمه به‌صورت زیر محاسبه می‌شود:

$$s(X,y)=\sum_{i=0}^{n}A_{y_i,y_{i+1}}+\sum_{i=1}^{n}P_{i,y_i} \qquad (1)$$

که در آن A ماتریسی مربعی از امتیازات انتقالی با اندازه k+2 است. Aij نشان‌دهنده امتیاز انتقال از برچسب iام به برچسب jام است. y_0 و y_n نشان‌دهنده برچسب آغازین و پایانی یک جمله هستند.

در فرآیند تشخیص موجودیت‌های اسمی به هریک از کلمات جمله یک برچسب[۱۶] تخصیص داده می‌شود. سیستم تشخیص موجودیت‌های اسمی معرفی شده در این مقاله برای اسامی شخص، مکان و سازمان به ترتیب از برچسب‌های LOC، PER و ORG استفاده می‌کند. علاوه بر این بعضی از موجودیت‌های اسمی خود از چندین کلمه تشکیل می‌شوند، بدین منظور از قالب برچسب‌گذاری IOB[۱۷] نیز استفاده می‌شود. B-Label کلمه ابتدایی یک موجودیت اسمی را نشان می‌دهد. I-Label سایر کلمات موجودیت اسمی به غیر از کلمه ابتدایی و O-Label کلمات خارج از موجودیت اسمی را نشان می‌دهند. در این مقاله از قالب IOBES استفاده شده‌است. که E-Label

15-Conditional Random Field

16-Label

17- IOB (Inside, Outside, Beginning)

کلمه انتهایی موجودیت اسمی و S-Label نشان‌دهنده این است که موجودیت اسمی تنها از یک کلمه تشکیل شده‌است.

۴- مدل پیشنهادی: نمایش برداری جامع[18] (AWR)

در این بخش به معرفی مدل پیشنهادی خود، نمایش برداری جامع می‌پردازیم. هدف از این مدل ارائه نمایشی از بردارهای کلمات به‌عنوان ورودی شبکه‌های عصبی LSTM است که بتوانند در تشخیص بهتر موجودیت‌های اسمی متن به‌کار روند. در این مدل به جای اینکه تنها از نمایش برداری در سطح نویسه به‌عنوان ورودی شبکه LSTM استفاده شود، نمایش برداری جامع‌تری برای هر یک از کلمات موجود در جمله می‌سازیم. بردار AWR از ترکیب بردارهای زیر ساخته می‌شود: ۱) نمایش برداری در سطح نویسه[19] (CLWE) ۲) بردارهای کلمات از پیش آموزش دیده[20] (PWE) ۳) نمایش برداری نحوی[21](SWE).

در ادامه روش ساخت هر یک از این بردارها و در نهایت نحوه ترکیب آن‌ها با یکدیگر شرح داده می‌شود.

۱,٤. بردار کلمات ورودی

مقادیر لایه‌های ورودی شبکه عصبی LSTM-CRF بردارهایی از کلمات مستقل هستند. این کلمات خود از نویسه‌هایی ساخته شده‌اند و به همین صورت قابل پردازش در شبکه‌های عصبی نیستند. به همین دلیل، یافتن نمایشی از کلمه[22] که مناسب این امر باشد، ضروری به نظر می‌رسد. نمایشی که برای این منظور به‌کار برده می‌شود، استفاده از بردار کلمات تعبیه شده[23] است. این بردارها با استفاده از روش‌های مختلفی که در ادامه شرح داده خواهند شد، به هر یک از کلمات موجود در واژگان زبان، یک نمایش عددی اختصاص می‌دهند.

۲,٤. نمایش برداری در سطح نویسه

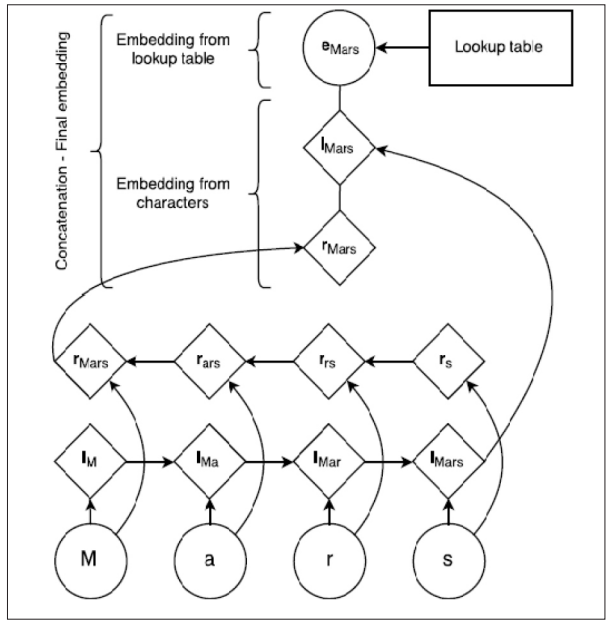
اولین بردار از سه بردار سازنده نمایش برداری جامع AWR که برای هر کلمه پیشنهاد می‌شود، نمایش برداری در سطح نویسه CLWE است. از آنجایی که در هر زبانی تعداد واژگان بسیار زیاد است و داده‌های آموزشی برای تشخیص موجودیت‌های اسمی محدود هستند، ساخت یک نمایش برداری خاص برای هر کلمه کار دشواری خواهد بود. از سویی دیگر در بعضی زبان‌ها مشاهده می‌شود که با عوض شدن املای کلمه، معنای آن و حتی اسم بودن یا نبودن آن نیز ممکن است تغییر یابد. بنابراین باید به دنبال نمایشی از کلمات بود که به املای آن‌ها حساس است. برای این منظور می‌توان برای هر کلمه از نمایش برداری نویسه‌های آن استفاده کرد [۲۷]. این کار باعث استخراج ویژگی‌هایی از کلمه می‌شود که در سطح نویسه هستند [۲۸].

یادگیری بردارهای تعبیه شده در سطح نویسه این مزیت را دارند که نمایش‌های برداری از کلمات را به‌دست می‌آوریم که مربوط به یک دامنه خاص هستند. برای مثال با استخراج نمایش‌های برداری کلمات از یک پیکره متنی پزشکی، مشاهده می‌شود اصطلاحات خاصی که هم ریشه هستند، بردارهای کلماتی مشابه یکدیگر خواهند داشت. از جمله روش‌های ساخت بردارهای کلمات با استفاده از نمایش برداری در سطح نویسه می‌توان به موارد زیر اشاره کرد: ۱- استفاده از شبکه عصبی Bi-LSTM ۲- استفاده از شبکه عصبی CNN . در ادامه به شرح هریک از روش‌های ذکر شده خواهیم پرداخت.

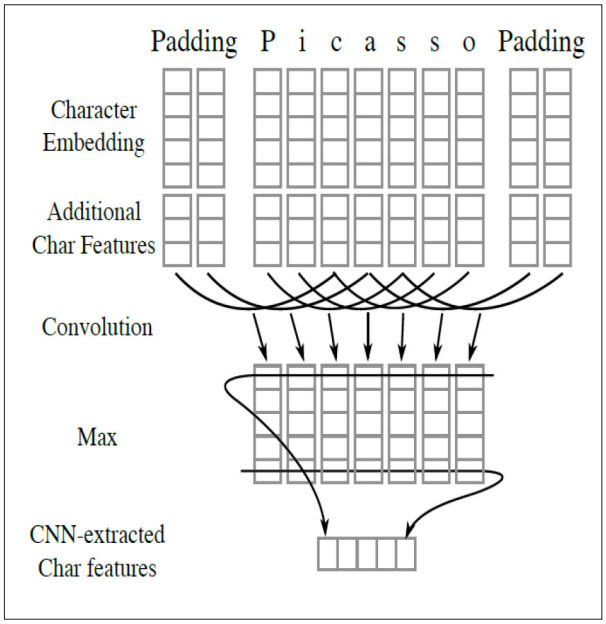
۱,۲,٤. استفاده از شبکه عصبی Bi-LSTM

در شکل شماره ۲ معماری شبکه عصبی Bi-LSTM برای ساخت بردار هر کلمه با استفاده از نویسه‌های آن ترسیم شده است. در ابتدا نویسه‌های کلمه مورد نظر از جدول جستجوی[24] نویسه استخراج می‌شوند. این جدول حاوی بردارهایی برای هر نویسه موجود در آن زبان است که به‌صورت تصادفی مقدارهای عددی به آن‌ها اختصاص داده شده است. سپس این نویسه‌ها به صورت مستقیم

18-Aggregative Word Representation (AWR)
19-Character Level Word Embedding
20-Pre trained Word Embedding
21- Syntactical Word Embedding
22-Word Representation
23-Word Embedding
24- Lookup Table



شکل ۲: معماری مدل شبکه Bi-LSTM برای ساخت بردار کلمات در سطح نویسه[۲۰] . نویسه‌های کلمه Mars از جدول جستجوی نویسه استخراج شده و یک به یک به LSTM دو جهته داده می‌شوند. خروجی این شبکه با بردار کلمه مربوط از جدول جستجوی کلمات ترکیب شده تا نمایش برداری نهایی برای کلمه Mars به‌دست بیاید.



شکل ۳: معماری مدل شبکه CNN برای ساخت بردار کلمات در سطح نویسه [۱۷] . بردار نویسه مورد نظر و بردار نوع آن پد گذاری شده و به ورودی شبکه عصبی CNN داده می‌شوند. سپس از طریق یک لایه کانوولوشن و یک لایه بیشینه بردار ویژگی مربوط به آن نویسه استخراج می‌شود.

و معکوس به ترتیب، به ورودی شبکه عصبی LSTM رو به جلو و LSTM رو به عقب داده می‌شوند. خروجی این دو شبکه با هم ترکیب شده و یک نمایش برداری اولیه برای کلمه به‌دست می‌آید. در نهایت این بردار با بردار کلمه مربوطه از جدول جستجوی کلمات[۲۵] ترکیب شده و بردار نهایی برای کلمه ساخته می‌شود. جدول جستجوی کلمات نیز شامل بردارهای تمام واژگان زبان می‌شود که از قبل آموزش دیده‌اند و مقادیر عددی آن‌ها به دست آمده است.

۲,۲,٤. استفاده از شبکه عصبی CNN

شبکه عصبی دیگری که برای ساخت بردار کلمه پیشنهاد شده است، شبکه‌های عصبی کانوولوشن می‌باشد (شکل شماره ۳). در این روش ابتدا برای هر نویسه یک بردار تعبیه شده و یک بردار نوع نویسه تشکیل می‌شود. مقادیر بردارهای تعبیه شده از طریق یک جدول جستجوی نویسه به‌دست می‌آیند. به این صورت که برای هر نویسه یه بردار با ابعاد ۲۵ در نظر گرفته شده که مقادیر هر یک از خانه‌های آن به‌طور تصادفی در بازه [۰/۵, ۰/۵-] مقداردهی شده‌اند. بردار نوع نویسه نیز یک بردار با ابعاد ٤ است که هر نویسه را با توجه به این که حرف بزرگ است یا کوچک یا جزو علائم نگارشی (نقطه، ویرگول و...) دسته‌بندی می‌کند، و به هر کدام از آن‌ها یک مقدار خاص نسبت می‌دهد. سپس هر یک از بردارهای کلمات با توجه به اندازهٔ پنجره CNN، از طریق نویسه‌های پدینگ خاصی، پد گذاری می‌شوند. در آخر این بردارها به ورودی شبکه عصبی کانوولوشن داده می‌شوند و با استفاده از یک لایه کانوولوشن و یک لایه بیشینه، بردار ویژگی برای هر نویسه به‌دست می‌آید.

۳,٤. بردارهای از پیش آموزش دیده

دومین بردار تشکیل دهنده روش پیشنهادی AWR، بردار کلمات از پیش آموزش دیده PWE است. مقادیر بردارهای از پیش آموزش دیده، از طریق آموزش و آزمایش بر روی پیکره‌های زبانی[۲۶] بزرگ به‌دست می‌آیند. از جمله این مجموعه بردارهای منتشر شده می‌توان به

25- Word Lookup Table

26- Corpora

جدول ۱: بردار موقعیت کلمه

کلمه	Jim	traveled	to	London	with	his	mother
موقعیت	۱	۲	۳	۴	۵	۶	۷

بردارهای استنفورد گلو[۲۷] که بر روی شش میلیارد کلمه از ویکی پدیا و متون وب آموزش دیده اشاره‌کرد. مورد دیگر بردارهای کلمه گوگل[۲۸] است که بر روی صد میلیارد کلمه از اخبار گوگل آموزش دیده‌اند.

٤,٤. نمایش برداری نحوی

سومین بردار تشکیل دهنده روش پیشنهادی AWR، نمایش برداری نحوی کلمه SWE است. این بردار نسبت به دو بردار دیگر CLWE و PWE از اهمیت بیشتری برخوردار است زیرا استفاده از اطلاعات نحوی کلمات در جمله مانند برچسب‌های اجزای سخن، نقش‌های وابستگی[۲۹] و جایگاه کلمه در جمله، کمک فراوانی در یافتن ویژگی‌هایی می‌کند که در تشخیص موجودیت‌های اسمی کاربرد خواهند داشت [۲۹]. بردار SWE خود از سه بردار زیر تشکیل می‌شود. ۱) بردار موقعیت کلمه ۲) بردار برچسب اجزای سخن و ۳) بردار وابستگی. در این مقاله ابتدا برای هر کلمه سه بردار ذکر شده ساخته می‌شود و سپس این بردارها با یکدیگر ترکیب شده و تحت عنوان بردار نحوی کلمه ارائه می‌شوند. در ادامه روش ساخت این سه بردار شرح داده می‌شود.

۱,٤,٤. بردار موقعیت کلمه

اولین برداری که برای ساخت بردار نحوی کلمه استفاده می‌شود، بردار موقعیت کلمه است بردارهای موقعیت کلمه[۳۰]، بردارهایی هستند که ترتیب قرارگیری کلمات در جمله را نشان می‌دهند. این بردارها از طریق محاسبه فاصله کلمه مورد نظر با سایر کلمات به‌دست می‌آیند؛ یا به هریک از کلمات بر اساس جایگاه آن‌ها در جمله یک شماره ترتیبی داده می‌شود (جدول ۱).

27- Stanford's GloVe Embedding: http://nlp.stanford.edu/projects/glove/
28- Google's word2vec Embeddings: https://code.google.com/p/word2vec/
29- Dependency Roles
30- Word Position Embedding

جدول ۲: بردارهای برچسب اجزای سخن

برچسب POS	بردار وان هات مربوط به هر برچسب				
فعل	۱	۰	۰	۰	۰
فاعل	۰	۰	۱	۰	۰
قید	۰	۱	۰	۰	۰
صفت	۰	۰	۰	۰	۱

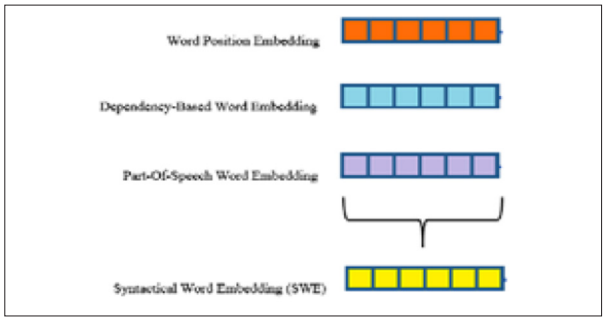
۲,٤,٤. بردار برچسب اجزای سخن[۳۱] (POS)

دومین برداری که برای ساخت بردار نحوی کلمه معرفی می‌شود، بردار برچسب اجزای سخن است. برای یافتن برچسب اجزای سخن از عمل برچسب‌گذاری اجزای سخن استفاده می‌شود که عمل انتساب برچسب به کلمات تشکیل دهنده یک متن است. این برچسب‌گذاری بر اساس نقش آن کلمه در متن مانند اسم، فعل، قید، صفت و غیره صورت می‌گیرد. برای تبدیل این برچسب‌ها به بردار از روش وان هات[۳۲] استفاده می‌شود که برای هر یک از برچسب‌ها یک بردار با مقادیر خاص تعریف می‌شود. در این روش به یکی از خانه‌های بردار مقدار یک و بقیه مقدار صفر در نظر گرفته می‌شود. این مقدار دهی به‌صورتی انجام می‌شود که هر بردار با دیگری متفاوت باشد (جدول ۲).

۳,٤,٤. بردار وابستگی

بردار سومی که برای ساخت بردار نحوی کلمه معرفی می‌شود، استفاده از بردار وابستگی است. بردارهای کلمات مبتنی بر وابستگی[۳۳] به بررسی روابط نحوی بین کلمات موجود در جمله می‌پردازد. از جمله این روابط نحوی می‌توان به اطلاع از انواع گروه‌های به‌کار رفته در جمله (گروه اسمی و گروه فعلی) و یا وابستگی کلمات با یکدیگر (فاعل، مفعول و قید) اشاره کرد [۳۰]. تفاوت این بردار،

31- Part-Of-Speech Tagging
32- One hot vector
33- Dependency-Based Word Embedding



شکل ۴: برداری نحوی کلمه

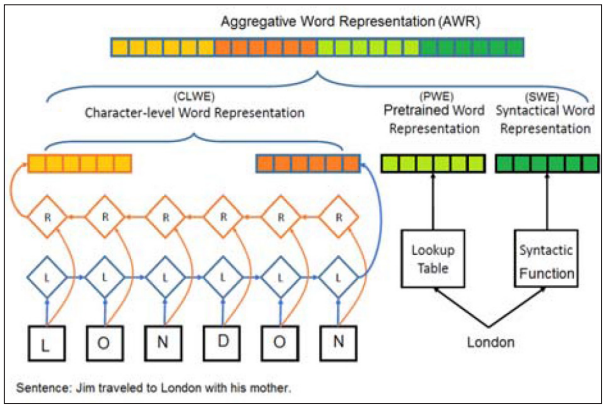
با بردار اجزای سخن در این است که برچسب نحوی کلمات در این روش با توجه به کل جمله استخراج می‌شود؛ بنابراین قادر است ارتباط معنایی کلماتی را که دور از یکدیگر هستند را نیز پیدا کند.

در روش پیشنهادی ما، بعد از محاسبه سه بردار موقعیت، POS و وابستگی برای هر کلمه، بردار نحوی کلمه از طریق جمع خانه‌های معادل این سه بردار ساخته می‌شود (شکل ٤).

٥,٤. معماری مدل AWR

در شکل شماره ٥ ساخت نمایش برداری جامع برای کلمه London نشان داده شده است. ابتدا نمایش برداری در سطح نویسه (بردار CLWE) از طریق شبکه عصبی Bi-LSTM (سمت چپ شکل) ساخته می‌شود. برای این کار، نویسه‌های کلمه London به‌صورت مستقیم (چپ به راست) و معکوس (راست به چپ) به ترتیب به ورودی شبکه عصبی LSTM رو به جلو و LSTM رو به عقب داده می‌شوند. خروجی این دو شبکه عصبی با هم ترکیب شده و بردار CLWE برای کلمه London به‌دست می‌آید.

سپس بردار از پیش آموزش دیده PWE کلمه London از جدول جستجوی کلمات استخراج می‌شود. بعد نمایش برداری نحوی کلمه London توسط تابع نحوی با توجه به جمله‌ای که در آن قرار گرفته (Jim traveled to London with his mother) به‌دست می‌آید. در نهایت از ترکیب این سه بردار به‌صورت متوالی نمایش برداری جامع (AWR) کلمه London ساخته می‌شود. همچنین در شکل شماره ٦ الگوریتم مدل پیشنهادی نیز بیان شده است.



شکل ۵: معماری مدل AWR . سمت چپ تصویر شبکه عصبی Bi-LSTM، تصویر وسط جدول جستجوی کلمات از پیش آموزش دیده و تصویر سمت راست تابع ساخت بردار نحوی را نمایش می‌دهد.

۵- آزمایش‌ها

در این بخش به بررسی جزئیات پیاده‌سازی مدل پیشنهادی خود، مانند تنظیم و مقداردهی اولیه پارامترهای مختلف می‌پردازیم.

۱,۵. معرفی مجموعه داده‌ها

مدل AWR بر روی مجموعه داده‌های CoNLL 2003 [۳۱] آموزش و آزمایش شده است که در آن موجودیت‌های اسمی مربوط به زبان انگلیسی و آلمانی برچسب‌گذاری شده‌اند. برای تهیه این مجموعه داده‌ها از مجموعه متون خبرگزاری رویترز برای زبان انگلیسی و مجموعه متون خبری روزنامه Frankfurter Rundschau برای زبان آلمانی استفاده شده است؛ سپس توسط عده‌ای از دانشگاه آنتورپ[۳٤] عمل برچسب‌گذاری برای کلمات انجام گرفته است. این مجموعه داده از ١٤٩٨٧ جمله که معادل ٢٠٣٦١ کلمه است، تشکیل شده است. ساختار مجموعه داده‌ها به این صورت است که در آن هرکلمه از جمله در یک سطر جداگانه قرار می‌گیرد و پس از پایان یافتن هر جمله یک خط خالی وجود دارد. هر سطر از چند بخش تشکیل شده‌است که اولین بخش کلمه مورد نظر، دومین بخش برچسب اجزای سخن مربوط به آن و آخرین بخش برچسب موجودیت اسمی کلمه (مکان، شخص، سازمان) است.

34- University of Antwerp

Algorithm: Aggregative Word Representation (AWR)

Input: Sentence
Normalization (Sentence)
Tokenization (Sentence)
for every **word** in Sentence **do**

CLWE = Bi-LSTM (word)
PWE = Lookup-Table (word)
SWE = Syntactical-function (word)
end for
AWE = compute-vector (CLWE, PWE, SWE)
Output: AWE

شکل۶: الگوریتم مدل AWR

۲،۵. مقدار دهی اولیه پارامترها

مدل LSTM-CRF به‌کار رفته در مدل ما، از یک لایه برای شبکه عصبی LSTM رو به جلو و از یک لایه برای LSTM رو به عقب استفاده می‌کند. ابعاد هریک از این لایه‌ها ۱۰۰ در نظر گرفته شده است. با توجه به آزمایش‌های انجام شده، کم و زیاد کردن ابعاد شبکه LSTM تاثیری در کارایی مدل ندارد. برای آموزش شبکه عصبی از الگوریتم بازگشت به عقب و گرادیان کاهشی تصادفی[۳۵] با نرخ یادگیری ۰،۰۱ استفاده شده است. نرخ دراپ آوت برابر ۰/۵ قرار داده شده است. بالا بردن این نرخ تاثیری منفی در نتایج آزمایش دارد و کم کردن آن باعث افزایش زمان آموزش مدل می‌شود.

برای نمایش برداری کلمات موجود در جدول جستجوی کلمات از بردارهای کلمات از پیش آموزش دیده استنفورد گلو استفاده شده است. کلماتی که در مجموعه مورد نظر نیستند، به‌صورت تصادفی در بازه $[-\sqrt{\frac{۳}{dim}}, +\sqrt{\frac{۳}{dim}}]$ مقداردهی می‌شوند که dim ابعاد بردار مورد نظر را نشان می‌دهد. برای مقداردهی تصادفی بردارهای نویسه موجود در جدول جستجوی نویسه‌ها نیز از این رابطه استفاده می‌شود.

۳،۵. تنظیم پارامترهای اصلی

ما ابعاد بردار کلمات را برابر $\{۲۰۰,۱۰۰,۵۰\}$ و ابعاد بردار نویسه‌ها را برابر $\{۲۵,۱۰ ,۵۰\}$ قرار دادیم. در نهایت ابعاد ۱۰۰ برای بردار کلمات و ۲۵ برای بردار نویسه‌ها انتخاب شدند. ابعاد بردارهای برچسب اجزای سخن و موقعیت کلمه و بردار وابستگی برابر ۵ در نظر گرفته شده است.

35- Stochastic gradient descent (SGD)

۶- ارزیابی مدل و نتایج

دراین بخش عملکرد مدل پیشنهادی AWR با مجموعه ویژگی‌های مختلف، با نتایج سایر مدل‌های منتشر شده مقایسه می‌شود. نتایج به‌دست آمده در جدول شماره ۳ در سه قسمت نمایش داده شده است. در قسمت اول نتایج مدل پیشنهادی ما، قسمت دوم نتایج مدل‌های مختلف شبکه‌های عصبی برای تشخیص موجودیت‌های اسمی و در قسمت سوم نتایج روش‌های وابسته به زبان در این حوزه ارائه شده است.

۱،۶. معیار ارزیابی F1

استفاده از معیار دقت[۳۶] در سیستم تشخیص موجودیت‌های اسمی، معیار دقیقی برای ارزیابی به شمار نمی‌رود زیرا رده‌های موجود در پیکره متنی متوازن نیستند. یک مدل امتیاز دهی که به طور استاندارد برای این منظور مورد تعریف شده است، استفاده از دو معیار صحت[۳۷] (P) و بازخوانی[۳۸] (R) است. تعریف آن‌ها به‌صورت زیر است:

36- Accuracy
37- Precision
38- Recall

$$Precision = \frac{T_p}{T_p + F_p} \qquad (۲)$$
$$Recall = \frac{T_p}{T_p + F_n}$$

که Tp پاسخ‌های مثبتی هستند که به درستی مثبت تشخیص داده شده‌اند، Fp پاسخ‌های اشتباهی هستند که به اشتباه درست تشخیص داده شده‌اند و Fn پاسخ‌های درستی هستند که سیستم به اشتباه، غلط تشخیص داده است. از ترکیب این دو معیار، معیاری به‌دست می‌آید که میانگین توافقی[۳۹] آن دو معیار است و به آن معیار F_1 می‌گویند[۳۲].

$$F۱ = \frac{۲PR}{P + R} \qquad (۳)$$

۲,۶. نتایج

در قسمت اول جدول شماره ۳ نتایج استفاده از ویژگی‌های مختلف مدل AWR یعنی استفاده از بردارهای در سطح نویسه، بردارهای از پیش آموزش دیده و بردارهای نحوی بر روی شبکه عصبی LSTM-CRF بررسی شده‌است. در ابتدا تنها از بردارهای PWE و CLWE به‌عنوان ویژگی‌های ورودی به شبکه LSTM برای تشخیص موجودیت‌های اسمی استفاده شد و به کارایی ۹۰/۵٤ دست یافتیم. سپس تاثیر استفاده از بردارهای برچسب اجزای سخن POS به عنوان یکی از ویژگی‌های نحوی کلمه بر روی مدل بررسی شد و کارایی آن برابر ۹۰/۸۹ گردید. در نهایت بردار نحوی کلمه SWE با بردارهای PWE و CLWE ترکیب شد و مشاهده گردید ساخت بردارهای کلمات با استفاده از هر سه این بردارها، کارایی مدل برابر ۹۱/۳۰ می‌شود. همچنین در روش‌های SWE+ Char(CNN) + PWE و SWE+ Char(Bi-LSTM) + PWE تاثیر ساخت بردارهای در سطح نویسه با دو شبکه عصبی مختلف CNN و Bi-LSTM بررسی شده و مشاهده گردید ساخت بردار کلمات در سطح نویسه به وسیله روش Bi-LSTM کارایی بالاتری نسبت به روش CNN دارد.

در قسمت دوم جدول شماره ۳ نتایج مدل‌های مختلف شبکه‌های عصبی بررسی گردیده است. بیشترین کارایی مربوط به چیو و نیکولز بوده (۹۱/٦۲) که از شبکه‌های عصبی کانوولوشن برای ساخت بردار کلمات در سطح نویسه به همراه بردارهای از پیش آموزش دیده و فرهنگ واژگان استفاده نموده‌اند [۱۷]. پس از آن معیار کارایی لمپل و همکاران (۹۰/۹٤) قرار می‌گیرد که از شبکه‌های عصبی LSTM برای بردار کلمات و بردارهای از پیش آموزش دیده شده استفاده نموده‌اند [۲۰]. کولبرت و همکاران [۱٤] و هانگ و همکاران [۱۵] به ترتیب به معیارهای کارایی ۸۹/۵۹ و ۹۰/۱۰ دست یافتند که روش پیشنهادی آن‌ها در بخش کارهای مرتبط توضیح داده شد.

در قسمت سوم جدول شماره ۳ نتایج مدل‌های مختلف وابسته به زبان، بررسی شده است. بهترین نتیجه در حوزه تشخیص موجودیت‌های اسمی حتی در مقایسه با بعضی از مدل‌های شبکه‌های عصبی، تاکنون مربوط به لو و همکاران [۳۳] بوده‌است که به معیار کارایی برابر ۹۱/۲۰، به وسیله استفاده مشترک از مدل تشخیص موجودیت اسمی و مدل پیوند موجودیت‌ها دست یافتند. همچنین مدل آن‌ها از ویژگی‌های دست سازی نظیر ویژگی‌های املایی، خوشه‌بندی وردنت[۴۰]، خوشه‌بندی براون[۴۱] و برچسب‌گذاری POS به همراه پایگاه‌های دانشی مانند فری بیس[۴۲] و ویکی‌پدیا استفاده می‌کنند. با این وجود همان‌طور که قبلا اشاره شد، استخراج این ویژگی‌های وابسته به زبان کار دشواری است و در تمامی زبان‌ها قابل پیاده‌سازی نیست.

در کل مدل پیشنهادی یعنی استفاده از نمایش جامع برداری AWR به عنوان ورودی شبکه عصبی LSTM-CRF عملکرد بهتری نسبت به سایر مدل‌های مطرح در حوزه تشخیص موجودیت‌های اسمی دارد. تنها در یک مورد کارایی روش مربوط به چیو و نیکولز به دلیل استفاده از فرهنگ‌های واژگان بیشتر شده‌است. ولی استفاده از فرهنگ‌های واژگان باعث وابستگی مدل به یک زبان خاص خواهد شد. مزیت مدل AWR این است

39- Harmonic mean

40-Word Net clusters
41-Brown clusters
42-Freebase

که علاوه بر انگلیسی در سایر زبان‌ها نیز قابل استفاده است.

۷- نتیجه گیری

در این مقاله یک نمایش جامع برداری برای استخراج ویژگی‌های کلمات به منظور تشخیص موجودیت‌های اسمی پیشنهاد شده است. این نمایش برداری از طریق ترکیب سه بردار زیر به‌دست آمده است: ۱) بردارهای کلمات در سطح نویسه ۲) بردارهای کلمات از پیش آموزش دیده و ۳) بردارهای نحوی. استفاده از نمایش برداری پیشنهادی علاوه بر افزایش کارایی در سیستم‌های شناسایی موجودیت‌های اسمی، باعث توسعه سیستم‌هایی می‌شود که کمترین وابستگی را به دامنه دارند؛ زیرا بزرگ‌ترین چالش حاضر در توسعه این سیستم‌ها وابستگی بسیار بالای آن‌ها به دامنه و پیکره متنی است که در آن توسعه داده شده‌اند. مقایسه نتایج مدل پیشنهادی با سایر مدل‌های ارائه شده در این حوزه نشان می‌دهد که استفاده از شبکه‌های عصبی آموزش دیده توسط ویژگی‌هایی در سطح نویسه و ویژگی‌های نحوی، می‌تواند برای شناسایی و طبقه‌بندی موجودیت‌های اسمی بسیار موثر و کارا باشد.

برای کارهای آتی پیشنهاد می‌شود از نمایش برداری جامع کلمات برای استخراج اطلاعات و تشخیص اسامی، در متون و پیام‌هایی به اشتراک گذاشته شده در شبکه‌های اجتماعی مختلف استفاده شود، زیرا اغلب اسامی موجود در این متون یا به‌صورت خلاصه شده هستند یا ساختاری غیر معمول دارند که با استفاده از ویژگی‌های وابسته به زبان نمی‌توان به آن‌ها دست پیدا کرد.

جدول ۳: مقایسه نتایج حاصل از مدل AWR با سایر مدل‌های تشخیص موجودیتهای اسمی

مدل‌ها		F1
مدل AWR	Char(Bi-LSTM)+ PWE	۹۰/۵۴
	POS + Char(Bi-LSTM) + PWE	۹۰/۸۹
	SWE+ Char(CNN) + PWE	۹۱/۰۹
	SWE + Char(Bi-LSTM) + PWE	۹۱/۰۳
مدل‌های شبکه‌های عصبی	Collobert et al. (2011) [14]	۸۹/۵۹
	Huang et al. (2015) [15]	۹۰/۱۰
	Lample et al. (2016) [20]	۹۰/۹۴
	Chiu and Nichols (2015) [17]	۹۱/۶۲
مدل‌های وابسته به زبان	Ratinov and Roth (2009) [13]	۹۰/۸۰
	Lin and Wu (2009) [11]	۹۰/۹۰
	Passos et al. (2014) [12]	۹۰/۹۰
	Luo et al. (2015) [33]	۹۱/۲۰

مراجع

1- Nadeau, D., Sekine, S. «A survey of named entity recognition and classification», Lingvisticae Investigationes 30(1), 3–26 (2007).

2-Yadav, V.and Bethard, S. " A survey on recent advances in named entity recognition from deep learning models" . In: Proceedings of the 27th International Conference on Computational Linguistics . Santa Fe, New Mexico, USA: Association for Computational Linguistics, 2145–2158, 2018.

3-Tahmoresnezhad, J., & Hashemi, S. "Visual domain adaptation via transfer feature learning". Knowledge and Information Systems, 50(2), 585-605, 2017.

۴-اصفهانی، راحتی قوچانی، جهانگیری، «سیستم شناسایی و طبقه‌بندی اسامی در متون فارسی»، فصلنامه پردازش علائم و داده‌ها، شماره۱۳،صفحات ۸۸-۷۷، ۱۳۸۹.

5- Crichton, G. , Pyysalo, S. , Chiu, B. , and Korhonen, A. "A neural network multi-task learning approach to biomedical named entity recognition". BMC Bioinformatics 18(1):368, 2017.

6-Goyal, A., Gupta, V. and Kumar, M. ,"Recent Named Entity Recognition and Classification techniques: A systematic review". Computer Science Review, Vol. 29, pp. 21–43, 2018.

7- Poostchi, H. , Zare Borzeshi, E. , Abdous, M. , Piccardi, M. "PersoNER: Persian named-entity recognition", Proceedings of the 26th International Conference on Computational Linguistics COLING'16, pp. 3381-3389, 2016.

8-Dernoncourt,F. , Young Lee,J. , Szolovits,P. , "Neuroner: an easy-to-use programfor named-entity recognition based on neural networks". arXiv preprint arXiv:1705.05487, 2017.

9-Tahmoresnezhad, J., & Hashemi, S , "A generalized kernel-based random k-samplesets method for transfer learning". Iranian Journal of Science and Technology Transactions of Electrical Engineering, 39(E2), 193-207, (2015).

10- Kuru,O. , Arkan Can,O. , Yuret.D. , "Charner: Character-level named entity recognition". In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 911–921, 2016.

11-Lin,D., Wu,X. "Phrase clustering for discriminative learn-

ing. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2, pages 1030–1038. Association for Computational Linguistics, 2009.
12- Passos,A., Kumar,V., McCallum, A. "Lexicon infused phrase embeddings for named entity resolution", arXiv preprint arXiv:1404.5367, 2014.
13-Ratinov, L., Roth, D., "Design challenges and misconceptions in named entity recognition" , In Proceedings of the Thirteenth Conference on Computational Natural Language Learning, pages 147–155. Association for Computational Linguistics. 2009.
14-Collobert,R., Weston,W., L´eon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. Natural language processing (almost) from scratch. The Journal of Machine Learning Research, 12:2493–2537, 2011.
15-Huang,Z., Xu,W., Yu,K.,"Bidirectional LSTM-CRF models for sequence tagging", CoRR, abs/1508.01991, 2015.
16- Zhou,J., Xu,W. "End-to-End learning of semantic role labeling using recurrent neural networks" , In Proceedings of the Annual Meeting of the Association for Computational Linguistics, 2015.
17-Chiu,J., Eric Nichols,E., "Named entity recognition with bidirectional lstm-cnns", arXiv preprint arXiv:1511.08308, 2015.
18-Ma, X., Hovy, E., "End-to-end sequence labeling via bidirectional lstm-cnns-crf " , In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers. 2016
19-Santos, C., Guimar˜aes, V., "Boosting named entity recognition with neural character embeddings. arXiv preprint arXiv:1505.05008, 2015.
20-Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., Dyer, C., "Neural architectures for named entity recognition". In NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016. Pages 260–270. 2016.
21-Ling, W., Chu-Cheng, L., Tsvetkov, Y., Amir, S., Astudillo, R., Dyer, C., Alan W Black, A., Trancoso. I.,
"Not all contexts are created equal: Better word representations with variable attention" , In Proc. EMNLP., 2015.
22-Bengio,Y., Simard,P., Frasconi , P., "Learning long-term dependencies with gradient descent is difficult" , Neural Networks, IEEE Transactions on, 5(2):157–166. 1994.
23-Hochreiter, S., Schmidhuber, J., "Long short-term memory.Neural Computation" , 9(8):1735–1780. 1997.
24-Graves, a., Schmidhuber, J., "Framewise phoneme classification with bidirectional LSTM networks" , In Proc. IJCNN, 2005.
25-Ling, W., Dyer, C., Black, A., Trancoso, I., Fermandez, R., Amir, S., Marujo, L., Tiago 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, pages 1520–1530. Association for Computational Linguistics. 2015.
26-Lafferty, J., McCallum, A., Pereira, F., "Conditional random fields: Probabilistic models for segmenting and labeling sequence data", In Proc. ICML. 2001.
27-Zhang,X., Zhao,J., LeCun, Y., "Character-level convolutional networks for text classification", In Advances in Neural Information Processing Systems, pages 649–657. 2015.
28-Kim, Y., Jernite, Y., Sontag, D., Rush, A., "Character-aware neural language models", CoRR, abs/1508.06615, 2015.
29- B. Y. Lin, F. Xu, Z. Luo, and K. Zhu, "Multi-channel bilstm-crf model for emerging named entity recognition in social media," in Proc. the 3rd Workshop on Noisy User-generated Text , pp. 160–165, 2017.
30-Komninos,A., Manandhar, S., "Dependency Based Embeddings for Sentence Classification Tasks", in Proc of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp 1490-1500, 2016.
31-Erik, F., Sang, T., Fien De Meulder, F., "Introduction to the conll-2003 shared task: Language independent named entity recognition", In Proc. CoNLL. 2003.
۳۲-رضائی شـریف آبادی، خسـروری زاده، «برچسـب زنی خودکار نقش‌های معنایی در جملات فارسـی به کمـک درخت‌های وابسـتگی»، فصلنامه پردازش علائم و داده‌ها، شماره ۲۷، صفحات ۳۸-۲۷،۱۳۹۵.
33-Luo, G., Huang, X., Lin, C., Nie, Z., "Joint named entity recognition and disambiguation" , In Proc. EMNLP. 2015.