مقاله علمی-ترویجی

مروری برالگوریتم حافظه کوتاه مدت طولانی و کاربردهای آن

میترا حسینی، گروه مهندسی کامپیوتر، واحد مشهد، دانشگاه آزاد اسلامی مشهد، مشهد، ایران، hoseinimitra@yahoo.com

مهرداد جلالی، گروه مهندسی کامپیوتر، واحد مشهد، دانشگاه آزاد اسلامی مشهد، مشهد، ایران، dr_mehrdadjalali@yahoo.com

چکیده

شبکه‌های عصبی بازگشتی مدلی از روش‌های یادگیری عمیق است که در چند دهه گذشته موضوع بسیار پرطرفداری بوده است. شبکه‌های عصبی LSTM روشی جدید از پردازش اطلاعات است که به‌خاطر ساختار زنجیرهای که دارد بر روی داده‌هایی با سری زمانی فعالیت می‌کند که منجر به رخدادهای قدرتمندی در حوزه یادگیری عمیق می‌شود. اگر LSTM ویژگی مهمی در دنباله ورودی در گام‌های ابتدایی را تشخیص دهد، به‌دلیل داشتن حافظه کوتاه مدت طولانی می‌تواند این اطلاعات را طی مسیر طولانی منتقل کند و این گونه وابستگی‌های بلندمدت احتمالی را دریافت و حفظ می‌کند. در این مقاله سعی شده است انواع LSTM و کاربرد آنها در پردازش متن در حوزه‌های طبقه‌بندی متون و تحلیل احساسات و استخراج ویژگی بررسی شود علاوه بر این ما شناسایی اهداف آینده را مد نظر قرار داده و مسیر تحقیق آینده برجسته شده است.

کلمات کلیدی: یادگیری عمیق، شبکه‌های بازگشتی LSTM، پردازش متن، استخراج ویژگی

A Survey of Long Short Term Memory (LSTM) a Lgorithm and its Application

Mitra Hoseini, Department of Computer Engineering, Islamic Azad University, Mashhad, Iran, hoseinimitra@yahoo.com

Mehrdad Jalali, Department of Computer Engineering, Islamic Azad University, Mashhad, Iran, dr_mehrdadjalali@yahoo.com

Abstract

Recursive neural networks are a model of deep learning techniques that has been a very popular topic for the past few decades. LSTM neural networks are a new way of processing information that, because of its chain structure, operates on time-series data that leads to powerful deep learning events if LSTM is an important feature in the input sequence. Recognize in the early stages because of its short-term memory, it can transmit this information over long distances and receive and maintain such potential long-term dependencies. In this article, we have tried to examine the types of LSTM and their application in text processing in the fields of text classification, emotion analysis and feature extraction. In addition, we have also considered identifying future goals and highlighting the path of future research

Keywords: Deep learning, Recurrent neural network LSTM, Natural language processing, Feature extraction

۱- مقدمه

شبکه‌های عصبی بازگشتی در سال ۱۹۸۰ ایجاد شدند اما در چند سال اخیر بوده است که این گونه از شبکه‌ها به‌طور گسترده مورد استفاده قرار گرفته‌اند. قدرت بازنمایی زیاد آنها به این دلیل است که قادر هستند محاسبات زیادی را به صورت موازی در چندین لایه انجام دهند و مدل مناسبی از داده‌ها ارایه دهند. این گونه از شبکه‌های عصبی به‌طور خاص برای پردازش داده‌های سری یا دنباله‌دار مفید هستند و در آنها هر نورون یا واحد پردازشی قادر به حفظ حالت داخلی به منظور حفظ اطلاعات مرتبط با ورودی قبلی می‌باشد. درشبکه‌های عصبی بازگشتی اطلاعات تنها از یک مسیر حرکت می‌کند که جهت آن رو به جلو می‌باشد.

درواقع اطلاعات باشروع از نورون‌های ورودی و گذر از لایه‌های پنهان (درصورت وجود) به سمت گره‌های خروجی می‌روند. در این شبکه حلقه یا دوری وجود ندارد[۱] از دیگر کاربردهای شبکه‌های عصبی بازگشتی در حوزه‌های مختلف می‌توان داده‌های مبتنی بر زمان یا سری‌های زمانی[۱] مثل داده‌های پیش‌بینی هوا[۲]داده‌های مالی و داده‌های دریافتی از سنسورها، ویدئوها، متن‌ها و...که در واقعیت به‌صورت سریال/سری[۲] هستند را نام برد. به‌طور کلی می‌توان گفت شبکه‌های عصبی بازگشتی خانواده‌ای از شبکه‌های عصبی‌اند که به طور ویژه جهت پردازش داده‌های سری(یا دنباله‌ها) طراحی شده‌اند که در ادامه به جزییات بیشتری پرداخته می‌شود. در بخش۳ درباره پیش زمینه‌ای از دلایل استفاده از شبکه عصبی بازگشتی و مشکلات شبکه‌های RNN بررسی می‌شود، در بخش ۴ به موضوع شبکه‌های LSTM و دلایل استفاده از آن وجزییات الگوریتم‌ها پرداخته می‌شود، در بخش ۵ در مورد شبکه‌های LSTM دو طرفه در حوزه تحلیل احساسات در متن پرداخته می‌شود، در بخش ۶ نوع دیگر از شبکه‌های LSTM که شبکه‌های GRU می‌باشد و کاربرد آنها در طبقه‌بندی متن مورد بررسی قرار می‌گیرد، بخش ۷ شامل بررسی LSTMهای درختی می‌باشد و فصل آخر جمع‌بندی و کارهای آینده است.

۲- پیش زمینه

۲-۱-کاربرد شبکه‌های عصبی

در شبکه‌های عصبی، بهبود عملکرد با توجه به یک حافظه بسیار طولانی مدت در پارامترهای مدل، محقق می‌شود. پس از یادگیری از مجموعه آموزش‌ها، یک شبکه عصبی هنگام نمایش مثال‌های مشابه، اما پیش‌تر دیده نشده، به احتمال زیاد تصمیم درستی می‌گیرد.

موارد بسیاری وجود دارد که داده‌ها به طور طبیعی توالی تشکیل می‌دهند و در این موارد، ترتیب و محتوا به همان اندازه مهم هستند. از جمله نمونه‌های داده‌های توالی ویدئو، موسیقی، توالی DNA و ... است. هنگام یادگیری از داده‌های توالی، حافظه کوتاه مدت برای پردازش داده‌ها بسیار کاربردی می‌باشند. برای همین مدتهاست که محققان به یادگیری ماشین و شبکه عصبی مکرر روی آورده‌اند.

- شبکه‌های عصبی بازگشتی RNN

RNN به‌طور کلی یک شبکه عصبی بازگشتی[۳] است. در این نوع شبکه‌ها چالش این است که حافظه کوتاه مدت محدود است چرا که آموزش شبکه‌های بسیار عمیق دشوار است و حافظهRNN بسیار کوتاه است.RNNها به دو نوع با توجه به ورودی تبدیل می‌شوند یا ورودی را مرحله به مرحله می‌بینند یا ورودی را به صورت یکجا می‌بینند[۳].

۲-۲-اصول محاسبات شبکه های عصبی

برای پردازش توالی از بردارهای x رابطه (۱) برای هرمرحله استفاده می‌شود.

$$h_t = f_w(h_{t-1}.x_t) \quad (۱)$$
$$h_{t=\tanh(w_{hh}h_{t-1}+w_{hx}x_t)}$$
$$y_{t=w_{hy}h_t}$$

که در آن h_t لایه پنهان جدید و f_w تابع با پارامترهای w و x_t ورودی برداری در هر مرحله و h_{t-1} ورودی لایه پنهان مرحله قبلی و w_{xh} وزنی که روی ورودی اعمال می‌شود و w_{hh}وزنی که روی لایه قبلی و w_{hy} وزنی که روی خروجی شبکه اعمال می‌شود وy_t خروجی است که متناسب با ورودی تشکیل می‌شود. نکته‌ای که در RNNها وجود دارد این است که پارامترها در هر مرحله ثابت هستند. منظور از ثابت بودن این نیست که وزن‌ها تغییر پیدا نمی‌کنند بلکه وزن یک پارامتر خاص در هر مرحله به‌روزرسانی می‌شود. وظیفه RNN این است که اطلاعات که پیش‌تر مشاهده شده را حفظ کنند و به کارهایی که در حال انجام است مرتبط سازند در بعضی مواقع ما نیاز داریم به اطلاعات گذشته نزدیک نگاه کنیم تا متوجه اطلاعات حال حاضر بشویم مانند زمانی‌که ما مدل زبانی‌ای ساخته‌ایم که تلاش می‌کنیم کلمه بعدی را با توجه به کلمات قبلی پیش‌بینی کنیم زمانی که فاصله بین کلمات کم است شبکه عصبی بازگشتی قابلیت پیش‌بینی دارند. ولی ممکن است مواردی وجود داشته باشدکه ما به اطلاعات بیشتری نیاز داشته باشیم برای تشخیص کلمه بعدی به صورت کلی ممکن است فاصله بین کلمات و جایی که به این اطلاعات نیاز داریم زیاد باشد[۴]. هرچه این فاصله افزایش پیدا می‌کند، شبکه‌های عصبی بازگشتی قدرت‌شان در به یادآوردن و استفاده از اطلاعاتی که در گذشته دورتر یاد گرفته‌اند کاهش پیدا می‌کند و به عبارتی توانایی استفاده از اطلاعات گذشته دورتر را ندارند. حافظه خوبی ندارند و نمی‌توانند اطلاعات را برای زمان طولانی نگهداری کنند به‌صورت تئوری شاید از عهده کار برآیند ولی در عمل این‌گونه نیست. این مشکل به‌دلیل وجود محو شدگی گرادیانت[۴] است[۵] که رابطه (۲) این موضوع را بیان می‌کند.

$$s_t = \tanh(U_{x_t} + W_{s_{t-1}})$$
$$\hat{y}_{t=softmax(v_{st})}$$
$$E_t(y_t.\hat{y}_t) = -y_t \log \hat{y}_t$$
$$E(y.\hat{y}) = \sum_t E_t(y_t.\hat{y}_t) \quad (۲)$$
$$= \sum_t y_{t \log \hat{y}_t}$$

U پارامتر وزن‌ها بین ورودی و سلول و V لایه پنهان y_tخروجی و به صورت Softmax است و s_tکه از U_xt و w_{st-1} اثر می‌پذیرد.

۲-۲-۱-تابع خطا

فرض کنید توالی کامل از کلمات داریم و در لایه پنهان مقداری کلمه تولید می‌شود با توجه به سلول‌های قبلی و ورودی در نهایت یک خروجی تولید می‌کنیم که بر اساس خروجی به‌دست آمده خطایی داریم، برای محاسبه V از رابطه (۳) استفاده می‌کنیم.

$$\frac{\partial E_3}{\partial v} = \frac{\partial E_3}{\partial \hat{y}_3}\frac{\partial \hat{y}_3}{\partial v} \quad (۳)$$
$$\frac{\partial E_3}{\partial y_3}\frac{\partial \hat{y}_3}{\partial z_3}\frac{\partial z_3}{\partial v}$$

$$z_3 = Vs_3$$

S1 به S2 وابسته است و S1 به S0 وابسته است و ما توالی بزرگ از ضرب‌ها داریم، در رابطه (۴) خواهیم دیدکه همین توالی ازضرب‌ها ما را دچار مشکل می‌کند.

$$\frac{\partial E_3}{\partial W} = \sum_{k=0}^{3} \frac{\partial E_3}{\partial \hat{y}_3}\frac{\partial \hat{y}_3}{\partial s_3}\frac{\partial s_3}{\partial s_k}\frac{\partial s_k}{\partial W} \quad (۴)$$

$$\frac{\partial s_3}{\partial s_1} = \frac{\partial s_3}{\partial s_2}\frac{\partial s_2}{\partial s_1}$$

$$\frac{\partial E_3}{\partial W} = \sum_{k=0}^{3} \frac{\partial E_3}{\partial \hat{y}_3}\frac{\partial \hat{y}_3}{\partial s_3}\left(\prod_{j=k+1}^{3} \frac{\partial s_j}{\partial s_{j-1}}\frac{\partial s_k}{\partial W}\right)$$

$$s_t = \tanh(Ux_t + Ws_{t-1})$$

برای تمام مسیرها که از W گذر می‌شود باید محاسبات را داشته باشیم. این ضرب‌های زیاد از W باعث به‌وجود آمدن مشکل می‌شود. اگر مقادیر W بزرگ باشد یک حالت انفجار به‌وجود می‌آورد و مقدار وزن‌ها به سمت بی‌نهایت می رود[۵]. زمانی‌که مشتقات در هم ضرب می‌شوند ممکن است به سمت صفر میل کنندکه به آن مشکل محوشدگی گرادیان[۶] گفته می‌شود. برای حل مشکل از LSTM استفاده می‌شود.

۳- LSTM (حافظه کوتاه مدت طولانی)

در LSTM سه گیت ورودی وجود دارد، گیت ورودی، اطلاعات از ورودی می‌گیرد، گیت خروجی، اطلاعات به خروجی می‌دهد، گیت فراموشی اطلاعات را از حافظه پاک می‌کند[۶] که ما به جزییات پیاده‌سازی آن نمی‌پردازیم و در این مقاله سعی شده کاربردهای آنها مد نظر نویسنده باشد.

LSTMها به انواع مختلفی تقسیم میشوند در این مقاله درباره مدل حافظه کوتاه مدت طولانی دوطرفه[۷]، دروازه های سلول های بازگشتی[۸]، حافظه کوتاه مدت طولانی درختی[۹] و کاربرد آنها در حوزه تحلیل احساسات، طبقه‌بندی متن و استخراج ویژگی و... بررسی شده است.

۴- Bidirectional Long hort-Term Memory (مدل حافظه کوتاه مدت طولانی دوطرفه)

LSTM دوطرفه یکی از روش‌های یادگیری عمیق است که برای یادگیری داده‌های ترتیبی[۱۰] مورد استفاده قرار می‌گیرد. LSTM دوطرفه به این صورت است که در زمان آموزش، یک بار از ابتدا به انتهای داده‌ها مورد بررسی قرار می‌گیرد و وزن‌ها و مقادیر بایاس اصلاح می‌شوند، یک بار از انتها به ابتدا فرآیند آموزش انجام می‌شود و ماتریس‌های مربوطه تنظیم می‌شوند و در انتها این دو گروه نورون، در کنار یکدیگر در قسمت خروجی قرار می‌گیرند. (به‌طور مثال اگر ۳۰۰ بعد داشته باشیم، خروجی این لایه ۶۰۰ بعد است که ۳۰۰ تای اول مربوط به LSTM از اول به آخر و ۳۰۰تای دوم مربوط به LSTM از آخر به اول است). به طور کلی از دو لایه حافظه‌ی کوتاه مدت طولانی تشکیل شده که به‌صورت موازی اجرا می‌شوند که در پردازش متن، لایه‌ی اول جمله و لایه‌ی دوم معکوس جمله را به‌عنوان ورودی دریافت می‌کند[۷].

- تحلیل احساسات

تحلیل احساسات مجموعه‌ای از روش‌ها، تکنیک‌ها و ابزارهای تشخیص و استخراج اطلاعات ذهنی مانند نگرش‌ها و نظرات از زبان است[۸]. این روش، پیام را با توجه به قطبیت آن که مثبت، منفی یا خنثی باشد، دسته‌بندی می‌کند. سه تکنیک متفاوت در این حوزه وجود دارد که عبارتند از: تکنیک یادگیری ماشین، تکنیک مبتنی بر قانون و تکنیک مبتنی بر فرهنگ واژه[۹] این موضوع یکی از فعال‌ترین زمینه‌های پژوهشی در پردازش زبان طبیعی و متن کاوی در سال‌های اخیر بوده است. تشخیص مثبت، منفی یا خنثی بودن نظر یا احساس بیان شده در ویژگی‌ها از داده‌های متنی شناسایی خودکار ویژگی‌ها، از طریق به‌کارگیری روش‌های نحوی مبتنی بر مدل‌سازی موضوعی[۱۱] و یادگیری عمیق[۱۲] امکان‌پذیر است[۱۰].

- طبقه‌بندی متون

طبقه‌بندی متن یکی از زیرمجموعه‌های داده کاوی است. در پردازش متن سعی می‌شود دانشی از متون خام استخراج شود. این موضوع از سال ۱۹۶۰ مورد توجه محققان بوده است و در دهه ۹۰ میلادی توجه به این موضوع بیشتر شده است. به دسته‌بندی متن، «برچسب‌گذاری متن» یا «طبقه‌بند متن» نیز گفته می‌شود[۲۷].

- LSTM های دوطرفه در تحلیل احساسات

معماری مقاله [۱۱] به شیوه‌ی تقسیم و غلبه ارایه شده است که ابتدا جملات را به انواع مختلفی تقسیم کرده سپس تحلیل احساسات را بر روی هر کدام از انواع جملات به‌طور مجزا انجام می‌دهد. هر گروه از جملات به یک شبکه عصبی یک بعدی برای رده‌بندی احساسات داده می‌شود.

رویکردهای یادگیری عمیق تا حدودی می‌توانند ویژگی‌هایی ساختاری و معنایی را از متن بدون نیاز به مهندسی ویژگی‌ها که دارای نیاز به کار و زمان بیشتری دارد، استخراج کنند. در سال ۲۰۱۴، کالچبرنو و همکارانش[۱۲] و کیم مدل‌های CNN مختلفی برای رده‌بندی احساسات ارایه دادند که در این مقاله [۱۳] به آن پرداخته شده است.

در این مقاله از شبکه‌ی d-CNN1 برای شناسایی قطبیت احساسات استفاده و از مدل BiLSTM-CRF برای استخراج هدف از جملات ورودی استفاده می‌شود و هر جمله طبق تعداد اهدافی که در آن به صورت آشکارا اظهار شده‌اند دسته‌بندی می‌شوند. BiLSTM-CRF از مدل‌های توالی عصبی عمیق هست که از یک لایه‌ی BiLSTM و یک لایه‌ی CRF که روی هم قرار گرفته‌اند که برای یادگیری توالی‌ها تشکیل شده است. به CRF[۱۳] مثل یک لایه نگاه می‌شود که مدل گرافیکی احتمالی غیرمستقیم است که توزیع‌های خطی ورود به سیستم بر روی خروجی‌های ساختار یافته مورد بررسی قرار می‌دهد.[۱۴]. در مدل BiLSTM-CRF از یک Softmax برای پیش‌بینی خروجی استفاده می‌شود. همچنین از تکنیک دوریز پس از لایه ورودی BiLSTM-CRF استفاده شده است تا از بیش برازش[۱۴] بر روی داده‌های آموزشی جلوگیری شود. این تکنیک توسط هی برای نخستین بار معرفی شد[۱۵]. استفاده از d-CNN 1 برای قطبیت احساسات بر روی هر نوع جمله شبکه d-CNN 1 ابتدا توسط کیم مطرح شد[۱۳].

این شبکه جملات با طول مختلف را به‌عنوان ورودی دریافت و بردارهای با طول ثابت را به‌عنوان خروجی تولید می‌کند. قبل از آموزش، بردارهای تعبیه شده[۱۵] برای هر کلمه از کلمات کل جملات ورودی تولید می‌شود. در لایه کانولوشن، چندین فیلتر با ابعاد پنجره‌های متفاوت بر روی بردارهای تعبیه شده حرکت می‌کنند تا یک کانولوشن یک بعدی انجام شود. از لایه‌ی Pooling، برای انتخاب کردن مفیدترین ویژگی‌های محلی به کار می‌رود. توابع فعالساز

غیرخطی element-wise اعمال می‌شود.
یکی از کاربردهای LSTM در این مقاله این است که می‌تواند چندین هدف که در یک جمله باشد و یا حتی اگر هدفی در جمله نباشد را بتواند به راحتی تشخیص دهد و طبقه‌بندی نماید.

– استفاده از LSTMدوطرفه در متون فارسی

با توجه به افزایش روز افزون داده‌های متنی و همچنین اهمیت اطلاعات استخراج شده از این داده‌ها، کاربرد طبقه‌بندی متون بسیار احساس می‌شود. نظرکاوی هم یکی از استفاده‌های طبقه‌بندی متون می‌باشد که کاربردهای آن در ارزیابی کالاها و استقبال جمعی از یک موضوع یا یک اتفاق بسیار به چشم می‌آید[۱۶]. تا به حال تمرکز بیشتر روش‌های ارایه شده بر روی زبان انگلیسی بوده است که در آن‌ها چالش‌های موجود در دیگر زبان‌ها مورد توجه قرار نگرفته است. یکی از عمده‌ترین چالش‌های موجود در زبان فارسی، تعداد زیاد پسوندها برای فرم‌های مختلف کلمات است مانند «من می‌روم» و «تو می‌روی»[۱۷] که با استفاده از LSTM سعی در حل این مشکل با توجه به فاعل کرده است. به‌طور تقریبی تا به حال بیشتر روش‌های معرفی شده در بحث نظرکاوی از روش‌ها و الگوریتم‌های سنتی یادگیری ماشین مثل رگرسیون لجستیک[۱۶]، بیزین ساده[۱۷] و ماشین بردار پشتیبان[۱۸] استفاده می‌کردند. که هر سه این الگوریتم‌ها با توجه به نحوه عملکردشان نیاز به بردار ویژگی‌ها با طول ثابت دارند. برای استخراج بردار ویژگی‌ها با طول ثابت، یک روش بسیار ساده و معقول استفاده از روش سبد واژگان است[۱۸] در این روش ساخت بردارها بدون توجه به مکان حضور کلمات در جمله و شباهت معنایی بین کلمات صورت می‌گیرد. در نظر نگرفتن این ترتیب کلمات باعث می‌شود، امکان تشخیص و تمیز دادن جملات با ترکیب کلمات یکسان و بار معنایی متفاوت وجود نداشته باشد[۱۹] که این موضوع باعث ایجاد خطا در روند انجام طبقه‌بندی می‌شود. با توجه به این امر که برای نظرکاوی در زبان فارسی دادگان بزرگ با ناظر وجود ندارد، این مقاله در ابتدا بردارهای کلمات خود را به‌صورت بدون ناظر بر روی تمام پیکره خود آموزش داده است. برای این امر از مدل چندتایی پرشی استفاده شده است مدل چندتایی پرشی سعی می‌کند با پیش‌بینی کلمات همسایه‌ی کلمه اصلی، یک نمایش برداری برای آن بدست بیاورد.[۲۰] پس از آن بردارها را به‌عنوان ورودی به مدل‌های شبکه عصبی مکرر و شبکه عصبی حافظه کوتاه مدت ماندگار دوسویه داده می‌شوند و مدل‌ها با استفاده از داده‌های برچسب‌دار آموزش می‌یابند. چون در عملیات طبقه‌بندی (نظرکاوی) جملات یکجا دریافت می‌شوند، امکان استفاده از دنباله‌های آینده نیز وجود دارد این نمایش برداری شامل اطلاعات سطح بالایی است که در آن خلاصه‌ای از اطلاعات متن موجود است. این نمایش برداری را می‌توان به‌عنوان ویژگی، در عملیات طبقه‌بندی استفاده کرد[۲۱].

– استفاده از LSTM دوطرفه و ترکیب با TF-IDF برای وزن‌دهی به احساسات

در مقاله [۲۲] برای نمایش کلمات از بردار کلمات استفاده می‌شود و برای وزن دهی به این بردار کلمات از الگوریتم سنتی TF-IDF بردارهای وزنی کلمه استفاده می‌شود. بردارهای کلمه‌ای وزندار به حافظه کوتاه مدت دو جهته BiLSTM وارد می‌شوند تا اطلاعات زمینه را به‌طور موثر درنظر بگیرند، گرایش احساسات توسط طبقه‌بندی شبکه عصبی پیشگام[۱۹] بدست می‌آید. در این مقاله، اینکه آیا کلمه‌ای حاوی اطلاعات که شامل احساسات است یا خیر، با تطبیق فرهنگ لغت احساسات تعیین می‌شود. در حال حاضر، فرهنگ لغت احساسات هاونت[۲۰]، فرهنگ ملی احساس دانشگاه ملی تایوان[۲۱] و فرهنگ لغت چینی[۲۲] و سه مورد هستند که به طور معمول استفاده می‌شوند. جملات توکن‌گذاری می‌شوند و بعد تبدیل به بردار و در مرحله بعد وارد شبکه عصبی بازگشتیBiLstm می‌شوند. برای کار تجزیه و تحلیل احساسات، قطبیت احساسات از فرمول که در مقاله [۲۲] وجود دارد استفاده می‌شود. تجزیه و تحلیل احساسات اساساً جز وظایف دسته بندی متن هستند و تنها با بردار کلمات نمی‌توان احساسات را بدست آورد. روند این مقاله این‌گونه است که در مرحله اول، وزن بردارهای کلمه که طبق فرمول TF-IDF به‌وجود می‌آید به عنوان ورودی‌های مدل BiLSTM استفاده می شود. سپس، بردارهای متن وارد طبقه‌بندی شبکه عصبی پیشخور می‌شوند عملکرد فعال‌سازی شبکه عصبی پیشخور عملکرد ReLU است.

– استفاده از LSTM دوطرفه و تابع خطا

در مقاله [۲۳] درباره سه مجموعه داده به‌طور کامل متداول مبتنی بر نظرات عمومی درباره فیلم‌ها صحبت شده است. از مجموعه داده‌ها مجموعه داده‌های MR [۲۴]، SST2[۲۵] و IMDb[۲۶] استفاده شده است. مشارکت‌های اصلی این مقاله عبارتند از: از روش تعبیه بدون نظارت استاندارد برای وزن‌های اولیه استفاده شده است، به نام بردار GloVe سپس ، یک لایه LSTM دو طرفه ، و یک لایه میانگین Pooling استفاده شده است. لازم به ذکر است، فقط به وابستگی‌های طولانی مدت میان کلمات در این مقاله توجه شده است. در این مطالعه، از یک روش باینری برای پیش‌بینی احساس مثبت یا منفی یک سند استفاده شده است. روش این مقاله شامل یک لایه ورودی که شامل یک لایه تعبیه شده، یک لایه BiLstm است و به دنبال آن گروهی از مجموعه لایه‌های متوسط و ماکسیمم Pooling و یک لایه سیگموئید در خروجی است. لایه ورودی به عنوان نقطه شروع شبکه در نظر گرفته می‌شود. لایه ورودی نمونه‌های داده را به‌عنوان دنباله‌ای از شاخص‌های منحصر به فرد با همان طول تبدیل می‌کند و شاخص‌ها با توجه با فرهنگ لغت به دست می‌آید. هر شاخص مربوط به یک کلمه منحصر به فرد در مجموعه داده‌ها به یک بردار ویژگی با ارزش واقعی تبدیل می‌شود. این بردارهای با ارزش واقعی در کنار هم قرار می‌گیرند و ماتریسی را تشکیل می‌دهند که ماتریس تعبیه شده نامیده می‌شود. از بردار Glove ۳۰۰ بعدی به‌عنوان یک بردار تعبیه کلمه از قبل آموزش دیده استفاده شده است. برای نمایش بهتر اطلاعات متنی، از مدل BiLstm استفاده می‌شود. خروجی‌های لایه BiLstm به‌طور همزمان به ماکسیمم و میانگین لایه‌های Pooling انتقال می‌یابد و بعد با هم ادغام می‌شوند و وارد لایه خروجی می‌شوند. از آنتروپی به‌عنوان یک تابع برای طبقه‌بندی احساسات باینری استفاده شده است که فرمول آن در مقاله [۲۳] وجود دارد. پیش پردازش با هدف حذف اطلاعات بیهوده از داده‌های متن خام انجام شده است و به‌عنوان یک کار اساسی در تجزیه و تحلیل احساسات در نظر گرفته می‌شود. شبکه‌های عصبی با استفاده از الگوریتم بهینه‌سازی شیب تصادفی SGD الگوهای اساسی و پیچیده را از داده‌ها می‌آموزند و در این مقاله نیز از SGD استفاده شده است. بهینه‌سازی فرآیندی برای جستجوی پارامترهایی است که می‌تواند یک عملکرد خاص را به بیشینه برساند یا آن را به کمینه برساند. جنبه مهم یادگیری عمیق، انتخاب یک تابع ضرر مناسب و سپس کاهش مقدار آن نزدیک به صفر است. در این مقاله از تابع ضرر آنتروپی باینری استفاده شده است که یک انتخاب ایده‌آل برای طبقه‌بندی باینری است. گره شبکه یک واحد محاسباتی است که از اتصال ورودی (های) وزنی، عملکرد انتقال و اتصال خروجی تشکیل شده است به طور خاص، اگر تعداد گره‌ها در یک لایه BiLstm بسیار

جدول ۱: مقایسه بین انواع روش‌های Bi-LSTM درحوزه تحلیل حساسات

مرجع	سال انتشار	الگوریتم	نوع شبکه عصبی	معیارهای اندازهگیری	مزایا
[۱۱]	۲۰۱۷	تحلیل احساسات	Bi-LSTM	دقت	ترکیبی از BI-LSTM و استفاده ازCRFکه نوعی مدل احتمالی است
[۲۷]	۲۰۱۷	تحلیل احساسات	Bi-LSTM	دقت-فراخوانی	به وجود آوردن نمایش برداری از متون فارسی
[۲۲]	۲۰۱۹	تحلیل احساسات	Bi-LSTM	دقت –فراخوانی	برای وزن احساسات فرمول خاصی درنظرگرفته می‌شود.
[۲۳]	۲۰۲۰	تحلیل احساسات	Bi-LSTM	صحت-دقت-هزینه محاسبات –زمان اجرا	استفاده از تابع ضرر برای کاهش خطا در BI-LSTM

کم باشد، توانایی یادگیری شبکه محدود می‌شود و ممکن است منجر به کمبود شود. در عوض، اگر تعداد گره‌ها خیلی زیاد باشد، پیچیدگی افزایش می‌یابد و ممکن است باعث بیش برازش شود برای این منظور، یک لایه BiLSTM را فقط با ۱۶ گره استفاده شده است و این مدل بدون هیچ‌گونه کم کاری یا پرتحرکی، عملکرد خوبی دارد. در لایه تعبیه، از مقدار دوریز ۰٫۳ استفاده شده است.

۵-Gated Recurrent Unit (دروازه‌های سلول‌های بازگشتی)

معماری GRU یا همان Gated Recurrent Unit در سال ۲۰۱۴ توسط چو و همکاران معرفی شد[۲۸] « دروازه‌های سلول‌های واحدهای بازگشتی» (Gated Recurrent Units Cells | GRU) نوعی از سلول‌های حافظه کوتاه مدت طولانی[۲۳] هستند. این سلول‌ها نیز از گیت‌هایی برای جلوگیری از مفقود شدن اطلاعات استفاده می‌کنند، اما این کار را تنها با استفاده از دو گیت «به روز رسانی»[۲۴] و «بازنشانی»[۲۵]انجام می‌دهند. این امر موجب می‌شود که آن‌ها اندکی کمتر شفاف باشند و در عین حال، سرعت آن‌ها اندکی بیشتر است زیرا همه جا از اتصالات کمتری استفاده می‌کنند. در اصل، دو تفاوت بین LSTM و GRU وجود دارد که در ادامه بیان شده‌اند. سلول‌های GRU، حالت سلول پنهانی که توسط یک گیت خروجی محافظت می‌شود را ندارند و ورودی را ترکیب می‌کنند و گیت را در به روزرسانی گیت، فراموش می‌کنند.

- ترکیبی از GRU و LSTM

در مقاله [۲۹] از دوتا LSTM استفاده شده است که به ترتیب و به طور همزمان اطلاعات زمینه را از طریق LSTM و GRU می‌گیرد. معماری آن این‌گونه است که از طریق محاسبه تشابه بین اطلاعات محتوایی که به‌طور موازی مورد بررسی قرار گرفته است ماتریسی مشترک ایجاد می‌کند. ورودی هر ماتریس یک بردار کلمه است به طور موازی اطلاعات وارد LSTM و GRU می‌شود، خروجی لایه LSTM که (SL) نامیده می‌شود و به‌عنوان سطر در نظر گرفته می‌شود و خروجی لایه GRU که (SG) نامیده می‌شود به‌عنوان ستون ماتریس است. ماتریس توجه تمام عناصر (SL)و (SG) را با هم مقایسه می‌کند. به دلیل افزایش تعداد پارامترها تعدادشان باعث بیش‌برازش[۲۶] می‌شود. از آنجا که LSTM و GRU اطلاعات متنی را گرفته‌اند، بردارهای ویژگی می‌توانند یک ماتریس متنی ایجاد کنند. به منظور استخراج اطلاعات معنی‌دار، مدل پیشنهادی ترکیب و جمع‌سازی را در ماتریس توجه انجام می‌دهد تا مشکل بیش برازش حل شود و به اطلاعات دقیق‌تر دست یابد.

- استفاده از GRU و متد Bag of Words در طبقه‌بندی متن

در مقاله [۳۰] به یکی از مشکلات در طبقه‌بندی متن پرداخته است این‌که چگونه می‌توان دقت طبقه‌بندی را افزایش داد و از مشکل خلوتی [۲۷] ، یعنی پراکندگی معناشناسی داده‌ها نسبت به زمینه در طبقه‌بندی متون جلوگیری کرد. در قسمت نمایش داکیومنت‌ها برای هر کلمه از روش Bag of Words استفاده شده است، این نمایش متنی وقوع کلمات را در یک سند تعریف می‌کند. برای جلوگیری از بیش برازش بهترین ویژگی‌ها به‌عنوان ورودی استفاده می‌شود. با توجه به مدل نمایش داکیومنت‌ها انتخاب ویژگی بسیار مفید و باعث افزایش دقت الگوریتم می‌شود.

- طبقه‌بندی براساس GRU و ماشین‌های پشتیبان

در مقاله[۳۱] از یک طبقه‌بندی متن براساس مدل مبتنی بر رویکردهای بهبود یافته با ارایه یک ماشین بردار پشتیبان خطیSVM به‌عنوان جایگزین Softmax در لایه خروجی نهایی یک مدل GRU استفاده شده است. علاوه بر این، تابع آنتروپی با یک تابع مبتنی بر حاشیه جایگزین می‌شود. ساختار کلی این مقاله این گونه است که در مرحله نخست ویژگی‌های مجموعه داده به شبکه GRU وارد می‌شوند و بعد مقادیر اولیه پارامترهای یادگیری وزن از طریق آموزش تنظیم می‌شوند. ارزش حالت سلول شبکه GRU بر اساس ویژگی‌های ورودی و پارامترهای یادگیری محاسبه می‌شود. برای به کمینه رساندن ضرر از روش بهینه‌سازی مادام استفاده شده است. در این مقاله به سه قسمت توجه می‌شود. نخست بر روی شبکهGRU متمرکز شده است که شیب ناپدید شدن و انفجار را به کم‌ترین مقدار برساند. درمرحله دوم از تابع فعال‌سازی واحد خطی اصلاح شده LRelu در معادله اصلاح شده جایگزین کرده است. ثابت شده است که واحدهای LRelu عملکرد بهتری نسبت به سیگموئید غیرخطی برای DNNهای پیش رو دارند و در آخر با اجرای ماشین بردار پشتیبان خطی SVM به‌عنوان جایگزینی Softmax سنتی در لایه خروجی نهایی یک مدل GRU سعی در بهبود عملکرد داشته است.

- شبکه عصبی Bi-GRU مبتنی بر توجه

این [۳۲] مقاله با استفاده از مکانیسم توجه، رویکرد جدیدی را برای طبقه‌بندی متن پیشنهاد می‌کند. در این مقاله یک شبکه عصبی Bi-GRU دو طرفه مبتنی بر توجه استفاده می‌شود که FABG در این مقاله نامیده می‌شود. از Bi-GRU برای یادگیری اطلاعات معنایی متن استفاده می‌کند. به منظور استفاده از مدل در اطلاعات کلیدی، از مکانیزم یانگ[۳] استفاده می‌شود. در [۳۳] مکانیسم توجه و شبکه‌های توجه سلسله مراتبی یانگ، تاثیر طبقه‌بندی متن را

جدول ۲:مقایسه بین انواع روش‌های طبقه‌بندی متن با شبکه‌های عصبی GRU

مرجع	سال انتشار	نوع الگوریتم	نوع شبکه عصبی بازگشتی	معیارهای اندازگیری	مزایا
[۲۹]	۲۰۲۰	طبقه بندی متن	GRU	دقت و سرعت همگرایی	ایجاد یک ماتریس خروجی مشترک از LSTM و GRU
[۳۰]	۲۰۱۹	طبقه بندی متن	GRU	دقت و میانگین مربعات خطا	استفاده از bag of word برای انتخاب ویژگی
[۳۱]	۲۰۲۰	طبقه بندی متن	GRU	دقت و صحت	استفاده از SVM به جای softmax
[۳۲]	۲۰۱۹	طبقه بندی متن	Bi-GRU	صحت	استفاده از لایه توجه در انتخاب کلمات کلیدی

بهبود می‌بخشد. یانگ فقط یک بار به کل خروجی‌های رمزگذار وزن اختصاص می‌دهد. به منظور توجه بیشتر به اطلاعات کلیدی از شبکه عصبی Bi-GRU دو طرفه بر اساس مکانیسم توجه استفاده می‌شود. این مکانیسم برای تمرکز بر اطلاعات کلیدی استفاده می‌شود FABG با اختصاص وزن به خروجی‌های رمزگذاری شده مراحل فعلی و قبلی که از مکانیسم توجه استفاده می‌کند، بردارهای هر مرحله را دوباره محاسبه می‌کند. با استفاده از تغییرات وزنی که در لایه توجه داده است جواب‌ها بهبود می‌یابند.

۶- Tree-LSTM (حافظه کوتاه مدت طولانی درختی)

نوع دیگراز LSTM نمایش درختی آن است در ابتدا LSTM به صورت ساختار زنجیره‌ای بوده است که برای افزایش قابلیت حفظ اطلاعات خیلی دور، به ساختار درختی توسعه داده شده است. مدل‌های غیرحساس به ترتیب، به‌دلیل عدم توانایی تشخیص تفاوت‌های معنایی و وابستگی‌های زبان طبیعی نسبت به مدل‌های حساس، به ترتیب کلمه یا ساختار نحوی، برای بازنمایی جملات مناسب نیستند. برای مثال مدل‌های غیرحساس به ترتیب، تفاوت بین دو جمله‌ی «گربه از درخت بالا می‌رود» و «درخت از گربه بالا می‌رود» را تشخیص نمی‌دهند. به همین دلیل است که به مدل‌های حساس به ترتیب متوالی و یا مدل‌های با ساختار درختی روی می‌آوریم. به ویژه مدل‌های با ساختار درختی به دلیل تفسیر نحوی ساختار جمله، یک گزینه مناسب و جذاب زبانی است. مدل‌های بازگشتی حساس به ترتیب، قادر نیستند که اطلاعات خیلی قدیمی را به یاد بیاورند بنابراین برای حل این مشکل مدل‌های درختی پیشنهاد شدند. مدل درختی ساختار نحوی جمله را نیز بررسی می‌کند پس برای بازنمایی عبارات مناسب‌تر است. مدل LSTM با ساختار درختی مانند LSTM استاندارد، دارای گیت‌های ورودی و خروجی و سلول حافظه و حالت مخفی می‌باشد. تفاوتLSTM استاندارد و LSTM درختی در این است که در مدل با ساختار درختی، بردارهای گیت شده و به روزرسانی سلول حافظه وابسته به حالت‌های تعداد فرزندان است و همچنین برای اینکه بشود اطلاعات را از هر فرزند دلخواه ترکیب کرد، به جای یک گیت فراموشی واحد، برای هر فرزند یک گیت فراموشی داریم. دو نوع درخت LSTM درختی داریم:

Child-Sum Tree-LSTMs، N-ary Tree-LSTM هر دو نوع قادرند که اطلاعات را از واحدهای چند فرزندی خود ترکیب کنند.

– مدل Child-Sum Tree-LSTMs

مانند LSTM است با این تفاوت که hj مجموع حالت‌های مخفی فرزندان از نخستین فرزند تا k امین فرزند می‌باشد و در بخش داخلی تعداد سلول حافظه و حالت مخفی و دروازه فراموشی به تعداد فرزندان است. همان‌طورکه در رابطه (۵) مشاهده می‌کنید.

$$
\begin{aligned}
h_j &= \sum_{k\in C_j} h_k \\
i_i &= \sigma(W_i x_j + U_i h_j + b_i) \\
f_{jk} &= \sigma(W_f x_j + U_f h_k + b_f) \\
o_j &= \sigma(W_o x_j + U_o h_j + b_o) \\
u_j &= tanh(W_u x_j + U_u h_j + b_u) \\
c_j &= i_j \odot u_j + \sum_{k\in C_j} f_{jk} \odot c_k \\
h_j &= o_j \odot tanh c_j
\end{aligned}
\tag{۵}
$$

که در آن دروازه ورودی، دروازه خروجی، سلول حافظه،لایه پنهان، گیت فراموشی برای هر فرزند، k بردار ورودی برای هرکلمه جمله مجموعه فرزند برای هر نود است.

– مدل LSTM درختی مرتب N تایی

اگر در یک LSTM درختی، فاکتور انشعاب بیشتر ازN باشد و یا اینکه فرزندان مرتب شده باشند، یعنی بتوان فرزندان را از ۱ تا n ایندکس کرد، میتوانیم از LSTM درختی مرتب N تایی استفاده کنیم. در این مدل ترتیب فرزندان مهم است پس با توجه به رابطه (۶) می‌توان نتیجه گرفت که

$$
\begin{aligned}
i_i &= \sigma(W_i x_j + \sum_{l=1}^{n} U_{il}\, h_{jl} + b_i) \\
f_{jk} &= \sigma(W_f x_j + \sum_{l=1}^{n} U_{fkl}\, h_{jl} + b_f) \\
o_j &= \sigma(W_o x_j + \sum_{l=1}^{n} U_{ol}\, h_{jl} + b_o) \\
u_j &= tanh(W_u x_j + \sum_{l=1}^{n} U_{ul}\, h_{jl} + b_u) \\
c_j &= i_j \odot u_j + \sum_{l=1}^{n} f_{jl}\, h_{jl} \odot c_{jl} h_j \\
h_j &= o_j \odot tanh c_j
\end{aligned}
\tag{۶}
$$

برای هر فرزند k یک ماتریس جداگانه در نظر گرفته می‌شود و باعث می‌شود شرایط بهتر فرزندان را در نظر بگیرد. به طور مثال یک حوزه انتخاباتی در نظر بگیرد که در آن فرزند چپ گره با یک عبارات اسمی مطابقت دارد و سمت راست فرزند با یک عبارات فعلی مطابقت دارد.

فرض کنیـد در این حالت تاکید بر عبارت فعل در نمایش سـودمند اسـت. سـپس می‌توان پارامترهای Ufk را به گونه‌ای آموزش داد که اجزای fj1 فرزندچپ نزدیک به ۰ (به عنوان مثال ، «فراموش») باشد، در حالـی که اجزای fj2 فرزند راسـت نزدیک به ۱ اسـت (به عنوان مثال، «حفظ») باشد[۳۴] .

– استخراج رابطه معنایی با استفاده از ساختار درختی LSTM

اسـتخراج رابطه معنایی برای ساخت خودکار گراف دانش بسیار مهم اسـت و از انـواع وظایف پردازش زبان طبیعی مانند پاسـخ به پرس و جو، جسـتجوی معنایی و پیامد پشـتیبانی می‌کند [۳۵]. وظیفه اسـتخراج رابطه معنایی[۲۸] بیشتر وظیفه شناسایی موجودیت از متون خام و اسـتخراج روابط معنایی بین موجودیت اسـتخراج شـده را بر عهـده دارد. روش‌های موجود فقط ویژگی‌های سـطح واژگان[۲۹] را در نظـر می‌گیرند و بیشـتر ویژگی‌های نحـوی [۳۰] را نادیده می‌گیرند و در نتیجـه عملکرد اسـتخراج رابطه ضعیفی دارنـد. در مقاله [۳۵] با اسـتفاده از حافظه کوتاه مدتLSTM، ویژگی‌های سـاختاری [۳۱]را بر اساس درخت وابستگی[۳۲] در یک جمله به وجود می‌آورد. برای افزایش عملکرد استخراج رابطه، LSTM متوالی دو طرفه با ساختار درختی با توجه به منظور شناسایی ویژگی‌های مبتنی بر کلمه [۳۳]و ارزش سهم کلمات در جمله اسـتفاده کرد. LSTM با ساختار درختی می‌تواند به راحتی روابط وابستگی را ترکیب کرده و اصطلاحات معنایی را دقیق‌تر کند تا روابط معنایی را بیشـتر تنظیم کند. ویژگی‌های استخراج شده هـر کلمه به LSTM متوالـی دو جهت داده می‌شـوند. خروجی این شبکه در هر مرحله زمانی ورودی لایه توجه است. با استفاده از جریان اطلاعات پایین به بالا، وابستگی‌های موجودیت با یکدیگر تعامل دارند تا نوع رابطه را به روشـی نظارت شده تعیین کنند. LSTM جهت بالا بـه پایین اطلاعات برتر را از گره پایین ترین جد مشـترک[۳۴]LCA به گره موجودیت[۳۵] در گره LCA منتقل می‌کند مسـیر LCA به عنوان یک درخت باینری و وابسـتگی نحوی در نظر گرفته می‌شود LSTM پاییـن به بالا و بالا به پایین در تعامل اطلاعاتی کار اسـتخراج رابطه نقش دارند که منظور همان LSTM دوطرفه با ساختار درختی است. سهم کلمات با توجه به روابطی که استخراج شده است متفاوت است LSTM. متوالی دو طرفه با مکانیسم توجه برای شناسایی ویژگی‌های کلمات در جمله شـامل اطلاعات موقعیتی موجودیت و سهم کلمات، که به‌عنوان ورودی‌های مجموعه LSTM ساختاری دو طرفه استفاده شده است. ویژگی های مبتنی بر کلمه و ویژگی‌های ساختاری مبتنی بر وابسـتگی نحوی برای بهینه‌سـازی عملکرد کار اسـتخراج رابطه بهم پیوسـته‌اند. پایین‌ترین لایه معماری لایه تعبیه اسـت که شامل بردارهای عددی[۳۶] هسـتند و بردار وابسـتگی[۳۷]، بردار برچسب‌گذاری کلمـات[۳۸]، درنمایشـی برداری به عنـوان ورودی LSTM دو طرفه استفاده می‌شود.

– استخراج رویداد باTree -LSTM

در این مقاله [۳۶] یک مدل استخراج رویداد جدید بر روی یک شبکه LSTM درختی و یک شبکه Bi-GRU ساخته شده است. در این مقاله به‌طور همزمان برای به دسـت آوردن نمایندگانی از جمله کاندیدهای رویدادها و شناسـایی نوع رویداد مورد بررسـی قرار می‌گیرد، که در مقایسـه بـا مواردی کـه از LSTM ، CNN یا فقط سـاختار درختی LSTM اسـتفاده می‌کند بهتر است. حالت پنهان هر گره در LSTM درختی برای پیش‌بینی برچسـب آرگومان های کاندید و شناسـایی/ طبقـه بندی تمام آرگومان‌های یک رویداد اسـتفاده می‌شـود. طبق استخراج خودکار محتوا[۳۹]، که اختصار ACE می‌نامند دو وظیفه برای اسـتخراج رویداد وجود دارد تشخیص رویداد و شناسایی نوع رویداد است. در روش سنتی نوع رویداد را با توجه به واژگان اصلی که توسط ACE داده شده است انتخاب می‌شود که نتیجه شناسایی نوع رویداد در روش بالا به دلیل مشـکلات آبشاری ناشـی از خطا در هر مرحله تاثیر زیادی بر عملکرد کلی سیسـتم اسـتخراج رویداد دارد. در عین حـال، اگر نتوان از اطلاعات عنصر رویـداد به‌عنوان مرجع در فرآیند شناسایی نوع رویداد اسـتفاده کرد، عملکرد تشخیص نوع رویداد نیز تاثیر منفی خواهد گذاشـت. دقت در تشـخیص کلمـه کاندید برای شناسـایی نوع رویداد با شناسایی کلمه اصلی بسیار مهم است وظیفه شناسایی نوع رویداد این است که، مدل به‌طور مستقیم نوع رویداد را با استفاده از Tree-LSTM و Bi-GRU تشخیص می‌دهد با استفاده از نمایشـی که از جمله به‌دست می‌آورد. این مدل از طریق ارایه دو نوع شـبکه، زمینه و اطلاعات نحوی[۴۰] از بازنمایی جمله واقعی را به‌دست می‌آورد. Tree-LSTM سـاختار درخت وابستگی هر جمله را به‌عنوان ورودی می‌گیرد تا اطلاعات نحوی به‌دسـت آید بـه تدریج اطلاعات کل درخـت را در هر گره قرار می دهد. جمله به‌عنوان ورودی Tree-LSTM در نظر گرفته می‌شـود که با استفاده از تعبیه جمله نمایشی غنی تر همراه با ویژگی‌های نحوی به‌دسـت آید. نمایش برداری یک کلمه در این مقاله از سه قسمت تشکیل شده است:جدول تعبیه‌سازی کلمه[۴۱]،تگ‌گذاری اجزای کلمه[۴۲]، انواع نهادها[۴۳].

۷- جمع‌بندی

حال که به بررسـی پردازش متن با اسـتفاده از انواع مختلف LSTM پرداخته ایم می‌توانیم بگوییم LSTM مدلی اسـت که، ساختار خوبی برای پیش‌بینی داده‌های متنی دارد. بسیاری از پیاده‌سازی‌های متنی دارای نمایش داده‌های متغیر در زمان‌های مختلف هستند که نیاز به رویکردهایی دارند که به دلیل سـازگاری آسـان آنها برای پیش‌بینی بسـیار مناسب اسـت از LSTM و مشتقات آن اسـتفاده شود.صنعت پـردازش متن با اسـتفاده از یادگیری عمیـق در زمینه‌های مختلفی کاربـرد دارد.تعداد کارهای تحقیقاتی هرسـاله افزایش می‌یابد با این حال ما تازه در سـال‌های ابتدایی این دوره هستیم، مطالعات بیشتری اجرا خواهند شـد. در این بررسـی ما می‌خواهیم تحقیقات پیشرفته شـبکه‌های عصبی بازگشتی در حوزه پردازش متن برجسته تر کنیم. تحقیقات ما نشـان می‌دهد روش‌های باورنکردنی در این زمینه وجود دارد. در این تحقیق ابتدا توضیحی از تاریخچه شبکه‌های عصبی گفته شده است، سپس مقدمه‌ای بر الگوریتم RNN و مشکلات و دلایل به وجود آمدن الگوریتم‌های LSTM مطرح شده است. بعد انواع LSTMها مورد بررسـی قرار گرفت و برای هر نـوع LSTM کاربردهای آن در حوزه های مختلفی مانند تشـخیص احساسـات، رویدادها، استخراج ویژگی و ... مقالاتی مورد بررسـی قرار گرفت و همچنین جزییات معمـاری و نقاط قوت و پارامترهای آن مقاله نیز مورد بررسـی قرار گرفت و برای هر بخش جدولی تشکیل شده و مشخصات مقاله در آن آورده شـده است و نشان دادیم که حل مساله با استفاده از LSTM و یا نوع خاصی از LSTM دقت را بهبود می‌بخشد.

– چالش‌ها و کارهای آینده

حال می‌خواهیم وضعیت فعلی تحقیق و دسـتاوردهای به‌دست آمده را بیـان کنیم. ما پیش‌تر به سـازگاری روش‌هـای LSTM در حوزه پردازش متن اشاره کرده‌ایم. تکنیک‌های LSTM بسیار امیدوارکننده هسـتند و بسـیار مفیدتر خواهد بود زمانیکه مشکلات این روشها در حوزه‌های مختلف مورد بررسـی قرار بگیرد. خلاصه‌سـازی روش‌های مختلف در جدول‌ها در بخش پایانی به یکپارچه‌سازی تجزیه و تحلیل در حـوزه های مختلف کمک می‌کنـد. همچنین به تازگی روش‌های

جدیدی برای بهبود پیش‌بینی‌های LSTM ارایه شـده اسـت و سعی براین بود در این روش‌ها ورودی‌های بهتری به LSTM داده شـود تا بتواند وابسـتگی‌های بین کلمات درست تشخیص داده شود تا بتواند به درستی وزن‌ها به‌روزرسانی شود و با توجه به کاربرد حوزه‌های آنها پیش‌بینی را بهینه‌سـازی کند. این تحقیق ممکن است دریچه جدید از فرصت‌هـا را در حوزه‌های مختلف با توجه بـه کاربردهای نام برده شده در پردازش متن باز کند مانند کاربرد در حوزه علم مواد در همه جنبه‌های شـیمی مواد، از کشف و طراحی مواد جدید تا مدل‌سازی و بهینه سازی عملکرد آنها، تعریف روابط ساختاری، ویژگی و شناسایی ویژگی‌های جدید را کشف خواهد کرد استفاده شود.

پی‌نوشت‌ها

1 Time series
2 Sequential
3 Recurrent Neural Networks
4 Vanishing Gradient
5 Exploding
6 Vanishing Gradiant
7 Bi-LSTM
8 GRU
9 Tree-LSTM
10 Sequential
11 Topic Modelling
12 Deep Learning
13 Conditional random fields
14 Overfitting
15 Word embedding
16 logistic regression
17 Naive Bayes
18 SVM
19 Feedforward neural networ
20 Hownet
21 National Taiwan University Sentiment Dictionary (NTUSD)
22 Dictionary of Tsinghua University
23 LSTM
24 Update
25 Reset
26 Overfit
27 Sparsity
28 Semantic relations
29 lexical-level feature
30 Syntactic feature
31 Structural features
32 Dependency tree
33 Word-based features
34 lowest common ancestor
35 Entity node
36 word embedding
37 Dependency embedding
38 Embedding pos
39 Automatic Content Extraction
40 Context and syntactic information
41 Word embedding table
42 PoS embedding table
43 Entity type

مراجع

[1] J. Heaton, I Goodfellow, Y Bengio, and A Courville, "Deep learning, Genetic Programming and Evolvable Machines", vol. 19, no. 1-2, pp. 305-307, 2017.
[2] Z. Zhang and Y. Dong, "Temperature Forecasting via Convolutional Recurrent Neural Networks Based on Time-Series Data," Complexity, vol. 2020, pp. 1-8, 2020.
[3] A. Amidi and S. Amidi. Recurrent Neural Networks cheatsheet. Available: https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-recurrent-neural-networks
[4] S. Hochreiter, Y. Bengio, P. Frasconi, and J. Schmidhuber, "Gradient flow in recurrent nets: the difficulty of learning long-term dependencies," ed.
[5] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," IEEE Transactions on Neural Networks, vol. 5, no. 2, pp. 157-166, 1994.
[6] C. Olah. (2015). Understanding LSTM Networks. Available: https://colah.github.io/posts/2015-08-Understanding-LSTMs/
[7] A. Graves, N. Jaitly, and A.-r. Mohamed, Hybrid speech recognition with Deep Bidirectional LSTM. 2013, pp. 273-278.
[8] M. V. Mäntylä, D. Graziotin, and M. Kuutila, "The evolution of sentiment analysis—A review of research topics, venues, and top cited papers," Computer Science Review, vol. 27, pp. 16-32, 2018/02/01/ 2018.
[9] M. D. Devika, C. Sunitha, and A. Ganesh, "Sentiment Analysis: A Comparative Study on Different Approaches," Procedia Computer Science, vol. 87, pp. 44-49, 2016
[10] F. Abid, M. Alam, M. Yasir, and C. Li, "Sentiment analysis through recurrent variants latterly on convolutional neural network of Twitter," Future Generation Computer Systems, vol. 95, pp. 292-308, 2019.
[11] T. Chen, R. Xu, Y. He, and X. Wang, "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN," Expert Systems with Applications, vol. 72, pp. 221-230, 2017/04/15/ 2017.
[12] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," arXiv preprint arXiv:1404.2188, 2014.
[13] Y. Kim, "Convolutional neural networks for sentence classification," arXiv preprint arXiv:1408.5882, 2014.
[14] A. McCallum, "Efficiently inducing features of conditional random fields," arXiv preprint arXiv:1212.2504, 2012.
[15] T. Duyu, Q. Bing, W. Furu, D. Li, L. Ting, and Z. Ming, "A Joint Segmentation and Classification Framework for Sentence Level Sentiment Classification," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 11, pp. 1750-1761, 2015.
[16] B. Liu, "Sentiment analysis and opinion mining," Synthesis lectures on human language technologies, vol. 5, no. 1, pp. 1-167, 2012.
[17] A. Bagheri, M. Saraee, and F. de Jong, "Sentiment classification in Persian: Introducing a mutual information-based method for feature selection," in 2013 21st Iranian conference on electrical engineering (ICEE), 2013, pp. 1-6: IEEE.
[18] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in International conference on machine learning, 2014, pp. 1188-1196.
[19] R. Socher, J. Pennington, E. H. Huang, A. Y. Ng, and C. D. Manning, "Semi-supervised recursive autoencoders for predicting sentiment distributions," in Proceedings of the 2011 conference on empirical methods in natural language processing, 2011, pp. 151-161.
[20] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in Advances in neural information processing systems, 2013, pp. 3111-3119.
[21] P. Wang, Y. Qian, F. K. Soong, L. He, and H. Zhao, "A unified tagging solution: Bidirectional lstm recurrent neural network with word embedding," arXiv preprint arXiv:1511.00215, 2015.
[22] G. Xu, Y. Meng, X. Qiu, Z. Yu, and X. Wu, "Sentiment analysis of comment texts based on BiLSTM," Ieee Access, vol. 7, pp. 51522-51532, 2019.
[23] Z. Hameed and B. Garcia-Zapirain, "Sentiment Classification Using a Single-Layered BiLSTM Model," IEEE Access,

ادامه در صفحه ۸۶

ادامه از صفحه ۷۷

vol. 8, pp. 73992-74001, 2020.
[24] B. Pang and L. Lee, "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales," arXiv preprint cs/0506075, 2005.
[25] R. Socher et al., "Recursive deep models for semantic compositionality over a sentiment treebank," in Proceedings of the 2013 conference on empirical methods in natural language processing, 2013, pp. 1631-1642.
[26] A. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies, 2011, pp. 142-150.
[27] B. Roshanfekr, S. Khadivi, and M. Rahmati, "Sentiment analysis using deep learning on Persian texts," in 2017 Iranian Conference on Electrical Engineering (ICEE), 2017, pp. 1503-1508: IEEE.
[28] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in NIPS 2014 Workshop on Deep Learning, December 2014, 2014, no. Conference contribution.
[29] S. Yu, D. Liu, W. Zhu, Y. Zhang, and S. Zhao, "Attention-based LSTM, GRU and CNN for short text classification," Journal of Intelligent & Fuzzy Systems, vol. 39, no. 1, pp. 333-340, 2020.
[30] M. Zulqarnain, R. Ghazali, M. G. Ghouse, and M. F. Mushtaq, "Efficient processing of GRU based on word embedding for text classification," JOIV: International Journal on Informatics Visualization, vol. 3, no. 4, pp. 377-383, 2019.
[31] M. Zulqarnain, R. Ghazali, Y. M. M. Hassim, and M. Rehan, "Text classification based on gated recurrent unit combines with support vector machine," International Journal of Electrical & Computer Engineering (2088-8708), vol. 10, 2020.
[32] Q. Tang, J. Chen, H. Lu, Y. Du, and K. Yang, "Full Attention-Based Bi-GRU Neural Network for News Text Classification," in 2019 IEEE 5th International Conference on Computer and Communications (ICCC), 2019, pp. 1970-1974: IEEE.
[33] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies, 2016, pp. 1480-1489.
[34] K. S. Tai, R. Socher, and C. D. Manning, "Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks," 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), no. Conference Proceedings, pp. 1556-1566, 2015.
[35] Z. Geng, G. Chen, Y. Han, G. Lu, and F. Li, "Semantic relation extraction using sequential and tree-structured LSTM with attention," Information Sciences, vol. 509, pp. 183-192, 2020.
[36] W. Yu, M. Yi, X. Huang, X. Yi, and Q. Yuan, "Make It Directly: Event Extraction Based on Tree-LSTM and Bi-GRU," IEEE Access, vol. 8, pp. 14344-14354, 2020.

***

com.
[21]M. R. Palattella , M. Dohler, A. Grieco, G. Rizzo, J. Torsner, T. Engel, and L. Ladid, “Internet of things in the 5G era: enablers, architecture, and business models,” IEEE Journal Sel. Areas Communication, vol. 34, no. 3, pp. 510-527, March 2016.
[22] L. Zhang, J. Liu, M. Xiao, G. Wu, Y.-C. Liang, and S. Li, “Performance analysis and optimization in downlink NOMA systems with cooperative full-duplex relaying,” IEEE J. Select. Areas Communication, vol. 35, no. 10, pp. 2398-2412, October 2017.
[23]L. Zhang, M. Xiao, G. Wu, M. Alam, Y.-C. Liang, and S. Li, “A survey of advanced techniques for spectrum sharing in 5G networks,” IEEE Wireless Communication, vol. 24, no. 5, pp. 44-51, October 2017.
[24]L. Zhang, Y.-C. Liang, and M. Xiao, “Spectrum sharing for internet of things: A survey,” IEEE Wireless Communication, Dol: 10.1109/MWC.2018.1800259, vol.26, no.3, pp.132-139, June 2019.
[25]M. Latva-aho, K.Leppänen ,”Key drivers and research challenges for 6G ubiquitous wireless intelligence (white paper), 6G Flagship, University of Oulu, Oulu Editor:, ISBN: 978-952-62-2353-7,Sep 2019.
[26] I. F. Akyildiz et al., “Combating the Distance Problem in the Millimeter Wave and Terahertz Frequency Bands,” IEEE Commun. Mag., vol. 56, no. 6, pp. 102–108, June 2018.
[27]P.Yang, Y.Xiao, M.Xiao and S.Li,” 6G Wireless Communications: Vision and Potential Techniques”, IEEE Network, Dol:10.1109/MNET.2019.1800418,pp.70-75,July 2019.
[28]I. Akyildiz, A.kak, A.Shuainie,” 6G and Beyond: The Future of Wireless Communications Systems”,IEEE Access, Dol: 10.1109/ACCESS. 2020.3010896, vol.8, pp. 133995-134030, July 2020.
[29]C-X. Wang and etal.”6G Wireless Channel Measurements and Models: Trends and Challenges”, Vehicular Technology Magazine, vol.15, no.4, pp.22-32, Dec.2020.
[30]M.Chowdhury, Md. Shahjalal, S.Ahmed, and Y.Min Jang,” 6G Wireless Communication Systems: Applications, Requirements, Technologies, Challenges, and Research Directions”, Open, vol.1, pp.955-975, July2020.
[31]S. Sekander et al., “Multi-Tier Drone Architecture for 5G/ B5G Cellular Networks: Challenges, Trends, and Prospects,” IEEE Communications Magazine, vol. 56, no. 3, pp. 96–103, Mar. 2018.
[32]P. Hajipour, A. Shahzadi, S. Ghazi-Maghrebi, “Interference mitigation for a heterogeneous satellite network based on spatial angle and transmission power”, International Journal of Electronics and Communications, pp.131-138, January 2019.
[33]P. Hajipour, A. Shahzadi, S. Ghazi-Maghrebi, “Improved performance for a heterogeneous satellite-cooperative network with best relay node selection”, China Communications, Vol. 16, No. 5, pp.93-105, June 2019.
[34H. Yao et al., “The Space-Terrestrial Integrated Network: An Overview,” IEEE Communications Magazine, vol. 56, no. 9, pp. 178–85, September 2018.
[35]J.Zhang, P. Tang, L.Yu, T. Jiang, L.Tian,”Channel measurements and models for 6G: current status and future outlook”, Frontiers of Information Technology & Electronic Engineering, vol.21, no.1, pp.39-61, March 2020.
[36]V.Petrov, J.Kokkoniemi, D.Moltchanov, J.Lehtomaki, Y.Koucheryavy, M.Juntti,“Last meter indoor terahertz wireless access:Performance insights and implementation roadmap,” IEEE Communications Magazine, vol. 56, no. 6, pp. 158–165, June 2018.
[37]Z. Chen, X. Ma, B. Zhang, Y. Zhang, Z. Niu, N. Kuang, W. Chen, L. Li, and S. Li, “A Survey on Terahertz Communications,” China Communications, vol. 16, no. 2, pp. 1–35, Feburary 2019.
[38]C.Han, Y.Wu, Z.Chen and X.Wang,” Terahertz Communications (TeraCom): Challenges and Impact on 6G Wireless Systems”, arXiv: 1912.06040v2 [eess.SP] 13, December 2019.
[39]T. W. Crowe, W. R. Deal, M. Schr¨oter, C.-K. C. Tzuang, and K. Wu, “Terahertz RF Electronics and System Integration,” Proceedings of the IEEE, vol. 105, no. 6, pp. 985–989, June 2017.
[40]J. M. Jornet and I. F. Akyildiz, “Graphene-based plasmonic nano antenna for terahertz band communication in nano networks,” IEEE Journal on Selected Areas in Communications, vol. 31, no. 12, pp. 685–694, December 2013, US Patent: 9,643,841 B2. [Online]. Available:http://www.acsu.buffalo.edu/_jmjornet/patents /2017/p1.pdf.
[41]ITU-R, 2015. IMT Traffic Estimates for the Years 2020 to 2030. Report M.2370, ITU-R, Geneva, Switzerland.
[42]CMRI, 2019. The Outlook and Demand Report for 2030+. China Mobile Research Institute, Beijing (in Chinese). https://cmri.chinamobile.com/news/5985.html [Accessed on Jan.4.2020].

***