وزارت ارتباطات و فناوری اطلاعات
شرکت ارتباطات زیرساخت

عنوان مقاله: الگوریتم های هوش مصنوعی و کشف اخبار جعلی در شبکه های اجتماعی

تهیه کننده:	مدرک و رشته تحصیلی	رشته شغلی	اداره کل/دفتر
مرتضی قاسمی	کارشناسی ارشد هوش مصنوعی و رباتیکز	کارشناس راهبری حراست	رئیس اداره حفاظت پرسنلی
حسین راهلی	کارشناسی ارشد علوم سیاسی	کارشناس راهبری حراست	مدیرکل حراست
محسن باقری	کارشناسی ارشد مدیریت دولتی	کارشناس راهبری حراست	معاون حفاظت پرسنلی

عنوان حوزه تحقیقاتی مورد نیاز شرکت: **شماره ردیف حوزه تحقیقاتی مورد نیاز شرکت** : ۲۰

جایگاه هوش مصنوعی (یادگیری عمیق، شبکه های عصبی و ....) در حوزه های کشف تقلب و امنیت

چکیده

جذابیت در دسترسی آسان، کم‌هزینه و سرعت بالای تبادل اطلاعات موجب گسترش استفاده از اینترنت، خصوصاً شبکه‌های اجتماعی در سطح جهانی شده است. در کنار مزایای فراوان این تکنولوژی‌ها، انتشار پیام‌های نامطلوب[1] در سطح اینترنت و شبکه‌های اجتماعی از جمله اخبار جعلی[2]، هرزنامه‌ها[3]، شایعات[4]، مطالب گمراه‌کننده، لینک های مخرب[5] و غیره، اثرات نامطلوب و مخرب بسیاری به بار می‌آورند و شناسایی و تشخیص به موقع آن‌ها حائز اهمیت می باشد؛ بنابراین تحقیق و بررسی در زمینه ایجاد روش‌ها و ابزارهای مختلف شناسایی پیام‌های نامطلوب بیش از گذشته توجه محققان را به خود جلب کرده است. ازآنجایی‌که الگوریتم‌های تولید و انتشار اخبار جعلی در شبکه‌های اجتماعی به‌صورت هوشمندانه[6] تغییر می‌کنند، بنابراین الگوریتم‌های شناسایی آن نیازمند تغییرات و بهبود مستمر می‌باشند. ازاین رو قصد داریم در این تحقیق با بهره‌گیری از الگوریتم‌های یادگیری ماشین[7] و الگوریتم‌های یادگیری عمیق[8] در حوزه پردازش زبان طبیعی[9]، به شناسایی بهتر اخبار جعلی در شبکه‌های اجتماعی بپردازیم.

کلیدواژه: اخبار جعلی، طبقه‌بندی، هرزنامه، یادگیری ماشین، یادگیری عمیق

مقدمه

با گسترش محبوبیت شبکه‌های اجتماعی[10]، امروزه شبکه‌های اجتماعی آنلاین به همه‌گیرترین وسیله ارتباطی و ابزار ا شتراک اطلاعات مبدل شده ا ست. کاربران پایه‌ای‌ترین جزء در سل سله‌مراتب شبکه‌های اجتماعی ه ستند و خود مسئول محتوایی هستند که به اشتراک می‌گذارند. جزء بعدی، اجتماع‌هایی است که بر پایه تعاملات کاربران و اطلاعاتی که به ا شتراک می‌گذارند، ایجاد می شود. این اطلاعات می‌تواند شامل لینک‌ها، عکس‌ها، ویدئوها، پیام‌ها و غیره باشد. از دیدگاه امنیت، شبکه‌های اجتماعی خصوصیات منحصربه‌فردی دارد از جمله اینکه دسترسی اطلاعات و تعاملات بر پایه اعتماد است و کاربران حجم قابل‌توجهی از اطلاعات شخصی خود را با دوستانشان به اشتراک می‌گذارند. متأسفانه شبکه‌های اجتماعی روش‌های قدرتمندی برای احراز هویت[11] فراهم نمی‌کنند و به‌راحتی می‌توان خود را جای شخص دیگری معرفی و بدون اجازه وارد شبکه مورد اعتماد یک کاربر شد. این مسئله حتی برای افراد مشهور حادتر است چون به‌راحتی هر درخواست دوستی‌ای را قبول می‌کنند. دیگر خاصیت شبکه‌های اجتماعی وجود

[1] Undesirable message
[2] Fake
[3] Spam
[4] Rumors
[5] Malicious links
[6] Intelligently
[7] Machine learning
[8] Deep learning
[9] Natural Language Processing
[10] Online Social Networks (OSN)
[11] Authentication

کاربران با سطوح متفاوت آگاهی نسبت به این تهدیدها ست[۱]. درحالی‌که اغلب کاربران نسبت به تهدیدهایی چون هرزنامه‌ها، اخبار جعلی، فیشینگ و غیره آگاهی دارند، نسبت به خطرات مخفی در شبکه‌های اجتماعی بی‌تفاوت هستند و به‌راحتی روی هر لینکی کلیک می‌کنند. این خاصیت شبکه‌های اجتماعی و همچنین تعداد زیاد کاربران عضو این شبکه‌ها، آن‌ها را به بهترین و محبوب‌ترین بستر برای فعالیت افراد سودجو مبدل نموده است[۱].
پیام‌های نامطلوب را می‌توان به‌صورت کلی در قالب اسپم در نظر گرفت و انتشار اخبار جعلی، شایعات، لینک‌های مخرب، وب‌سایت‌های تبلیغاتی و غیره را به‌عنوان پیام های نامطلوب در حوزه فعالیت اسپمرها برشمرد[۲]. هدف اصلی این تحقیق به‌صورت کلی استفاده از الگوریتم های هوش مصنوعی جهت شناسایی پیام‌های نامطلوب در شبکه‌های اجتماعی می‌با شد و شنا سایی اخبار جعلی که نمونه بارزی از پیام‌های نامطلوب می‌با شد به صورت عمیق مورد پژوهش قرار می‌گیرد بنابراین در ادامه ضمن بیان اهمیت موضوع، به معرفی مفهوم خبر جعلی، اهمیت خبر جعلی و تعریف مفاهیم م شابه خبر جعلی پرداخته خواهد شد. نمونه ای از ویژگی‌های[۱۲] مختلف پیام‌های نامطلوب و ار سال‌کننده‌های این پیام‌ها که به‌عنوان ورودی الگوریتم‌های مختلف یادگیری ما شین و یادگیری عمیق جهت طبقه بندی[۱۳] و تفکیک پیام‌های نامطلوب از پیام‌های مطلوب استفاده می‌شود معرفی می شوند؛ ویژگی‌های استخراج شده از محتوای متون خبری، شامل ویژگی‌های زبان‌شناسی[۱۴]، بصری و زمینه اجتماعی[۱۵] شامل ویژگی‌های مبتنی بر کاربر، مبتنی بر پست و مبتنی بر شبکه معرفی خواهند شد. فرآیند شناسایی اخبار جعلی و ویژگی‌های مهم در شناسایی آن موردبحث قرار خواهد گرفت. یادگیری ماشین، یادگیری عمیق و الگوریتم های آن ها، کاربردها و مجموعه داده‌های[۱۶] عمومی حاوی اخبار جعلی با ویژگی‌های آن ارائه خواهد شد. در انتهای این تحقیق به بیان نتیجه‌گیری و پیشنهاد برای پژوهش‌های آتی می‌پردازیم.

۱- خبر جعلی

خبر جعلی به قطعه خبری گفته می‌شود که به‌صورت عامدانه غلط و مورد قبول واقع شده است[۲]

۲- اهمیت خبر جعلی

شبکه‌های اجتماعی همانند چاقوی دولبه برای مصرف اخبار می‌باشند. شبکه‌های اجتماعی ازیک‌طرف با تسهیل د ستیابی به اخبار با هزینه کم و انت شار سریع اطلاعات، باعث ترغیب افراد به ج ستجو و م صرف اخبار می شود و از طرف دیگر موجب پخش گسترده اخبار جعلی می‌شوند که این اخبار برای گمراهی اذهان خوانندگان و حاوی اطلاعات غلط و با کیفیت پایین می‌با شند. انت شار اخبار جعلی به صورت گ سترده و با اهداف گوناگون از جمله منافع سیاسی، اقتصادی و غیره انجام می‌شود که می‌تواند اثرات مخرب و منفی بر افراد و اجتماع داشته باشد. برای مثال در انتخابات ریاست‌جمهوری سال ۲۰۱۶ آمریکا، تخمین زده شد بیش از یک میلیون توئیت[۱۷] مربوط به اخبار جعلی "Pizzagate"[۱۸] بوده است و ازاین‌رو پدیده "خبر جعلی" به‌عنوان کلمه سال در فرهنگ لغت مکوآری[۱۹] در سال

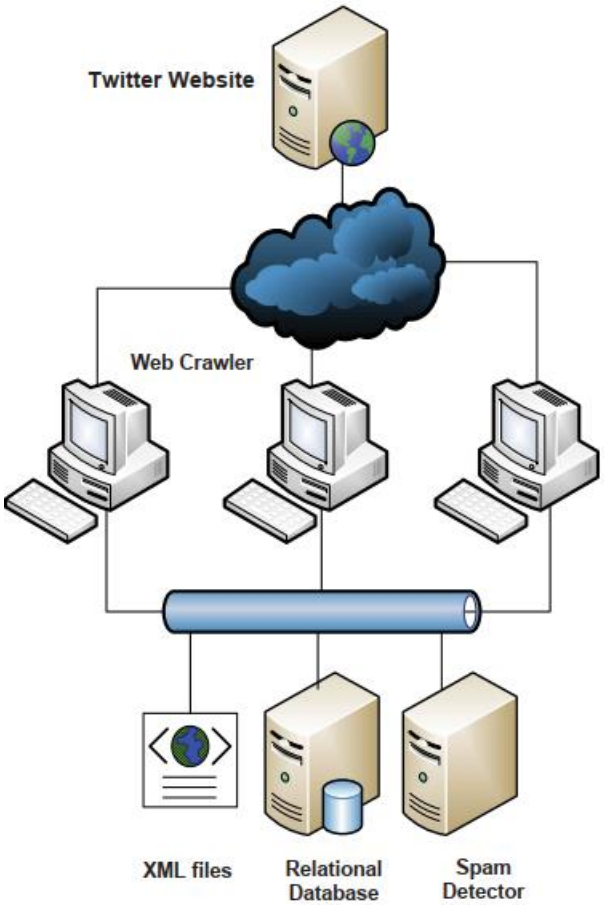
[12] Features
[13] Classification
[14] Features of linguistics
[15] Social context
[16] Datasets
[17] Tweet
[18] https://en.wikipedia.org/wiki/Pizzagate_conspiracy_theory
[19] Macquarie's Dictionary

۲۰۱۶ نام‌گذاری شد؛ بنابراین امروزه شناسایی اخبار جعلی از جمله تحقیقات ضروری می‌باشد که توجه بسیاری از محققان را به خود جلب کرده است[۳].

۳- فرآیند شناسایی پیام‌های نامطلوب

روش‌های تشخیص پیام‌های نامطلوب از سه بخش تشکیل شده‌اند که در شکل ۱ نشان‌داده‌شده است [۳, ۴]:

- بخش اول استخراج اطلاعات کاربران توسط کروالرها[۲۰].
- بخش دوم شناسایی مشخصه‌هایی است که به تشخیص پیام‌های نامطلوب کمک می‌کنند. به‌عنوان‌مثال، حساب‌های[۲۱] تولیدکننده پیام‌های نامطلوب معمولاً تعداد دنبال‌کننده‌های[۲۲] بیشتری به نسبت تعداد دنبال شونده[۲۳] دارند؛ بنابراین این نسبت می‌تواند یکی از ویژگی‌های تمیزکننده این گروه باشد.
- بخش سوم استفاده از الگوریتمی است که پیام‌های نامطلوب را به‌طور خودکار تشخیص دهد.



شکل ۱: ساختار کلی سیستم‌های تشخیص پیام نامطلوب[۴]

[20] Crawlers
[21] Accounts
[22] Followers
[23] Following

۴- استخراج ویژگی‌های محتوای خبر

ویژگی‌های محتوای خبری، به اطلاعات متا[۲۴] از یک قطعه خبری مربوط می‌شـــوند و بر اســـاس این اطلاعات خام، بازنمایی‌های[۲۵] مختلفی از ویژگی‌ها از جمله ویژگی‌های مبتنی بر زبان‌شـــناســـی[۲۶] و ویژگی‌های مبتنی بر بینایی[۲۷] ساخته می‌شوند. اطلاعات متا شامل موارد ذیل می‌باشد۵]:

- منبع خبر[۲۸]: نویسنده یا منتشرکننده قطعه خبری
- عنوان خبر[۲۹]: متن کوتاه در تیتر خبر با هدف جلب‌توجه خوانندگان که موضوع اصلی خبر را توصیف می‌کند.
- بدنه خبر[۳۰]: متن اصلی که جزئیات خبر را بیان می‌کند و گویای زاویه ناشر خبر می‌باشد.
- عکس و ویدئو: بخشی از بدنه محتوای یک قطعه خبری می‌باشد که اشاره به قالب داستان خبر دارد.

۵- نمونه ویژگی‌های استخراج شده از دیتاست های شبکه های اجتماعی توئیتر و فیس بوک

چند نمونه از ویژگی‌های استخراج شده از شبکه‌های اجتماعی توئیتر و فیس‌بوک مطابق جدول زیر می‌باشد:

جدول ۱ : نمونه اول از ویژگی‌های استخراج شده از شبکه‌های اجتماعی[۷]

Feature No.	Feature Name	Description
f1	account_age	The age (days) of an account since its creation until the time of sending the most recent tweet
f2	no_follower	The number of followers of this twitter user
f3	no_following	The number of followings/friends of this twitter user
f4	no_userfavourites	The number of favourites this twitter user received
f5	no_lists	The number of lists this twitter user added
f6	no_tweets	The number of tweets this twitter user sent
f7	no_retweets	The number of retweets this tweet
f8	no_hashtag	The number of hashtags included in this tweet
f9	no_usermention	The number of user mentions included in this tweet
f10	no_urls	The number of URLs included in this tweet
f11	no_char	The number of characters in this tweet
f12	no_digits	The number of digits in this tweet

[24] Meta information
[25] Representations
[26] Linguistics
[27] Vision
[28] News source
[29] News title
[30] News body

جدول ۲ : نمونه اول از ویژگی‌های استخراج شده از شبکه‌های اجتماعی[۷]

Index	Category	Feature	Comments
1	Tweet based	Source	Tweeting tools
2	Tweet based	Type	There are 4 types of tweets: Regular, Replies, Mentions and Retweets.
3	Tweet based	Retweet_count	The number of times the tweet is retweeted.
4	Tweet based	Favorite_count	The number of times the tweet is favorited
5	Tweet based	Hashtags_count	The number of hashtags appearing in the tweet.
6	Tweet based	Urls_count	The number of urls appearing in the tweet.
7	Tweet based	Mentions_count	The number of mentions appearing in the tweet.
8	Tweet based	Media_count	The number of media appearing in the tweet.
9	Tweet based	Symbols_count	The number of symbols (also called cashtag) appearing in the tweet.
10	Tweet based	Possibly_sensitive	If the tweet is regarded as possibly sensitive by Twitter
11	Profile based	Location	If the location field of profile is null.
12	Profile based	URL	If the URL field of profile is null.
13	Profile based	Description	If the description field of profile is null.
14	Profile based	Verified	If the user is verified by Twitter.
15	Profile based	Ff_ratio	Followers_count / Friends_count
16	Profile based	Followers_count	The number of followers of the user.
17	Profile based	Friends_count	The number of friends of the user.
18	Profile based	Statuses_count	The number of statuses the user post.
19	Profile based	Favourites_count	The number of tweets the user favorite.
20	Profile based	Listed_count	The number of lists the user create.
21	Profile based	Account_age	The lifespan of the account in terms of days.
22	Profile based	Default_profile	If the user is using a default profile.
23	Profile based	Default_profile_image	If the user is using a default avatar.

۶- دیتاست های معیار[۳۱] برای شناسایی اخبار جعلی

اگرچه توافقی بر روی دیتاست های معیار برای مسئله شناسایی اخبار جعلی وجود ندارد اما تعدادی دیتاست که در دسترس هستند به شرح ذیل می‌باشند[۸]:

- BuzzFeedNews: این مجموعه داده شامل یک نمونه کامل از اخبار منتشر شده در فیس‌بوک می‌باشد. این اخبار از ۹ آژانس خبری و در هفته آخر سال ۲۰۱۶ مربوط به انتخابات ایالات متحده آمریکا در تاریخ‌های ۱۹ الی ۲۳ و ۲۶ الی ۲۷ سپتامبر منتشر شده است و شامل ۱۶۲۷ قطعه خبر می‌باشد که ۸۲۶ قطعه مربوط به جریانات اصلی و ۳۶۵ قطعه مربوط به جناح چپ و ۵۴۵ قطعه مربوط به جناح راست می‌باشد.
- LIAR: این مجموعه داده از وبسایت PolitiFact جمع‌آوری‌شده است و شامل ۱۲۸۳۶ قطعه خبر می‌باشد که توسط انسان برچسب‌گذاری و از متن‌های مختلف مانند انتشار اخبار، مصاحبه‌های تلویزیونی یا رادیویی، سخنرانی‌های تبلیغاتی و غیره نمونه‌برداری[۳۲] شده است.
- BS Detector: این مجموعه داده از یک افزونه مرورگر[۳۳] به نام BS که برای بررسی صحت اخبار می‌باشد جمع‌آوری‌شده است. طریقه کار به این صورت است که منابع تمام لینک‌های موجود داخل صفحات وب را بررسی می‌کند تا آنهایی که مطابق لیست دامنه‌های اخبار جعلی هستند را تشخیص دهد. خروجی این افزونه، برچسب‌های مشخص‌کننده اخبار جعلی یا اخبار عادی می‌باشد.
- CREADBANK: این مجموعه داده شامل مقیاس گسترده‌ای از مجموعه داده اجماع منابع با تعداد تقریبی ۶۰ میلیون توئیت از ابتدای اکتبر ۲۰۱۵ به مدت ۹۶ روز می‌باشد. تمام توئیت ها به ۱۰۰۰ وقایع خبری شکسته شده اند که هر واقعه[۳۴] توسط ۳۰ مفسر[۳۵] از شرکت Amazon Mechnical Turk ارزیابی می‌شود.

[31] Benchmark datasets
[32] Sampling
[33] Browser plugin
[34] Event
[35] Interpreter

جدول۳: مقایسه ای از مجموعه داده‌های اخبار جعلی[۹]

Features / Dataset	News Content		Social Context		
	Linguistic	Visual	User	Post	Network
BuzzFeedNews	✓				
LIAR	✓				
BS Detector	✓				
CREDBANK	✓		✓	✓	✓

همان‌طور که مشـــاهده می‌شـــود هیچ‌کدام از مجموعه داده‌های اخبار جعلی، شـــامل تمام ویژگی‌های مبتنی بر زبان‌شناسی و مبتنی بر زمینه اجتماعی نمی‌شوند.

۷- یادگیری ماشین

در یادگیری ماشین، اسناد به سه روش طبقه‌بندی می‌شوند[۱۰]:

- **نظارت شده**[۳۶]: در این روش کلیه اسناد دارای برچسب[۳۷] می‌باشند که این برچسب می‌تواند توسط افراد خبره[۳۸] ایجاد شده باشد و اصطلاح طبقه‌بندی برای آن مرسوم می‌باشد. الگوریتم‌های این روش شامل بیزین، درخت تصمیم، k-نزدیک‌ترین همسایه، ماشین بردار پشتیبان، شبکه‌های عصبی و غیره می‌باشند.
- **غیرنظارت شده**[۳۹]: در این روش اسناد فاقد برچسب بوده و اصطلاح خوشه‌بندی[۴۰] برای آن مرسوم می‌باشد.
- **نیمه نظارتی**[۴۱]: در این روش تعدادی از اسناد دارای برچسب و تعدادی فاقد برچسب می‌باشند.

۸- یادگیری عمیق

حوزه یادگیری ماشـــین شـــاهد دوران طلایی خود اســـت زیرا یادگیری عمیق به‌آرامی در این حوزه پیشـــرو می‌شـــود. یادگیری عمیق از چندین لایه برای نمایش انتزاعات داده[۴۲] برای ســـاخت مدل‌های محاســـباتی اســـتفاده می‌کند. ازآنجایی‌که یادگیری عمیق بســـیار ســـریع رشـــد می‌کند، شـــبکه‌ها و معماری‌های جدید فراوانی هرچند ماه تولید می‌شوند. [۱۱]

۹- شبکه‌های عمیق و کاربردهای آن

نمونه‌هایی از شـــبکه‌های یادگیری عمیق (جدول ۴)، چارچوب‌های[۴۳] مختلف(جدول ۵)، روش‌های محبوب یادگیری عمیق در حوزه پردازش زبان‌های طبیعی(جدول ۶) و کاربرد هر یک از این شبکه‌ها(شکل ۲) به‌صورت مختصر معرفی شده اند.[۱۲]

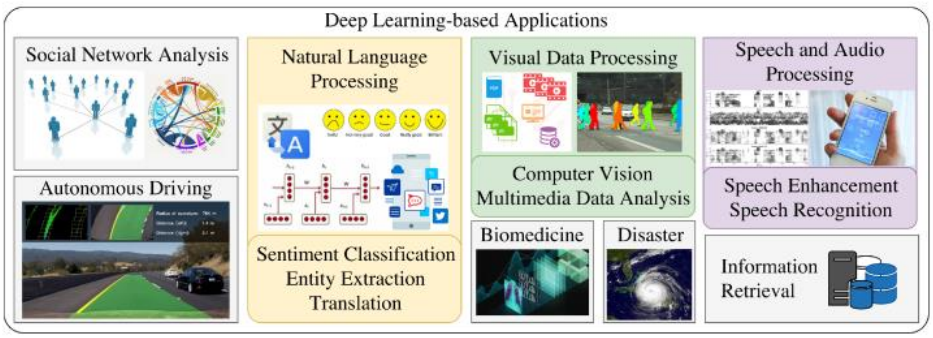
[36] Supervised
[37] Label
[38] Expert people
[39] Unsupervised
[40] Clustering
[41] Semi supervised
[42] Data abstractions
[43] Frameworks

جدول ۴: انواع شبکه‌های یادگیری عمیق[۱۲]

DL Networks	Descriptive Key Points	Papers
RvNN	Uses a tree-like structure Preferred for NLP	Goller et al. 1996 [47], Socher et al. 2011 [146]
RNN	Good for sequential information Preferred for NLP & speech processing	Cho et al. 2014 [20], Li et al. 2015 [93]
CNN	Originally for image recognition Extended for NLP, speech processing, and computer vision	LeCunn et al. 1995 [89], Krizhevsky et al. 2012 [86], Kim 2014 [79], Abdel-Hamid et al. 2014 [2]
DBN	Unsupervised learning Directed connections	Hinton 2009 [61], Hinton et al. 2012 [60]
DBM	Unsupervised learning Composite model of RBMs Undirected connections	Salakhutdinov et al. 2009 [135], Salakhutdinov et al. 2012 [136]
GAN	Unsupervised learning Game-theoretical framework	Goodfellow et al. 2014 [49], Radford et al. 2015 [130]
VAE	Unsupervised learning Probabilistic graphical model	Kingma et al. 2013 [81]

جدول ۵: مقایسه چارچوب‌های مختلف یادگیری عمیق[۱۳]

Framework	License	Core Language	Interface Support	CNN & RNN Support	DBN Support
Caffe [72]	BSD	C++	Python & MATLAB	Yes	No
DL4j [143]	Apache 2.0	Java	Java, Scala, & Python	Yes	Yes
Torch [25]	BSD	C & Lua	C/C++, Lua, & Python	Yes	Yes
Neon [69]	Apache 2.0	Python	Python	Yes	Yes
Theano [5]	BSD	Python	Python	Yes	Yes
MXNet [17]	Apache 2.0	C++	C++, Python, R, Scala, Perl, Julia, etc.	Yes	Yes
TensorFlow [1]	Apache 2.0	C++ & Python	Python, C/C++, Java, & Go	Yes	Yes
CNTK [173]	MIT	C++	Python, C++, & BrainScript	Yes	No



شکل۲: تعدادی از کاربردهای معمول یادگیری عمیق[۱۴]

جدول۶: روش‌های محبوب یادگیری عمیق در حوزه پردازش زبان‌های طبیعی[۱۵]

Paper	NLP Tasks	Architecture	Datasets
Socher et al. 2013 [147]	Sentiment Analysis	RNTN	SST
Kim 2014 [79]	Sentiment Analysis, General Classification	CNN	SST
Wehrmann et al. 2017 [164]	Sentiment Analysis	Conv-Char-S	MTD
Bahdanau et al. 2014 [9]	Translation	Bidir RNN Encoder-Decoder	WMT-14-EF
Cho et al. 2014 [20]	Translation	RNN Encoder-Decoder	WMT-14-EF
Wu et al. 2016 [166]	Translation	GNMT	WMT-14-EF WMT-14-EG
Socher et al. 2011 [145]	Paraphrase Identification	Unfolding RAE	MSRP
Yin et al. 2015 [172]	Paraphrase Identification, Question & Answer	ABCNN	WikiQA MSRP
Kågebäck et al. 2014 [75]	Summarization	Unfolding RAE	OD
Dong et al. 2015 [33]	Question & Answer	MCCNN	WQ
Feng et al. 2015 [39]	Question & Answer	CNN	IQA

۱۰- الگوریتم های کلاسیک یادگیری ماشین[۴۴]

انواع مدل‌ها یا الگوریتم‌های کلاسیک یادگیری ماشین وجود دارد که در ادامه به معرفی پرکاربردترین آن‌ها می‌پردازیم.

- الگوریتم ساده بیزین[۴۵]
- K نزدیک‌ترین همسایه[۴۶]
- رگرسیون لجستیک[۴۷]
- جنگل تصادفی[۴۸]
- درخت تصمیم[۴۹]
- ماشین بردار پشتیبان[۵۰]

۱۱- مدل‌های مبتنی بر شبکه عصبی و یادگیری عمیق[۵۱]

شبکه‌های عصبی عمیق[۵۲]، انقلابی در زمینه پردازش زبان طبیعی ایجاد کرده‌اند. شبکه‌های عصبی پیچشی[۵۳] و شبکه‌های عصبی بازگشتی[۵۴]، دو نوع اصلی از معماری شبکه‌های عصبی عمیق هستند که به طور گسترده‌ای برای

[44] Traditional Machine Learning Models
[45] Naïve bayes
[46] K-NN
[47] Logistic Regression
[48] Random forest
[49] Decision Tree
[50] Support vector machine
[51] Neural Network-Based and Deep Learning Models
[52] Deep neural networks (DNNs)
[53] Convolutional neural network (CNN)
[54] Recurrent neural network (RNN)

انجام وظایف مختلف در حوزه پردازش زبان طبیعی مورد استفاده واقع شده‌اند. شبکه‌های عصبی پیچشی در استخراج ویژگی‌های مکانی ثابت و شبکه‌های عصبی بازگشتی در مدل‌سازی واحدهای دنباله‌ای به‌خوبی عمل می‌کنند[۱۶].

۱۲- ارزیابی مدل های کلاسیک یادگیری ماشین و یادگیری عمیق بر روی دیتاست های اخبار جعلی

در این مقاله[۱۸]، بر روی دو مجموعه داده که از مخزن یادگیری ماشین[۵۵] و قابل‌دسترس در سایت Kaggle می باشد استفاده شده است. در مجموعه داده اول، هر مقاله خبری مطابق جدول۷ از چهار جزء تشکیل شده است و اندازه واژگان این مجموعه داده حدود ۲۹ مگابایت می‌باشد.

جدول ۷: مشخصات مربوط به مجموعه داده اول[۱۹]

Attribute	Type
# (id)	Numeric
title	Text
text	Text
label	Text (either REAL or FAKE)

در مجموعه داده دوم نیز هر مقاله خبری مطابق جدول۸ از چهار جزء تشکیل شده است و اندازه واژگان این مجموعه داده حدود ۱۲ مگابایت می‌باشد.

جدول۸: مشخصات مربوط به مجموعه داده دوم[۱۹]

Attribute	Type
URLs	Text
Headline	Text
Body	Text
Label	Numeric (either 1 for REAL or 0 for FAKE)

مجموعه داده اول به‌صورت ۶۰:۲۰:۲۰ برای آموزش[۵۶]، اعتبارسنجی[۵۷] و آزمون[۵۸] تقسیم شده است و شامل ۱۰۴۳۰۸ توکن منحصربه‌فرد[۵۹] و مجموع ۴۰۰۰۰۰ بردار کلمه در شبکه از پیش پردازش شده GloVe می‌شود. به‌صورت مشابه، ارزیابی مدل بر روی مجموعه داده دوم، دارای شرایط مجموعه داده اول و با تفاوت ۵۸۳۲۴ توکن منحصربه‌فرد انجام شده است.

نتایج ارزیابی برای مجموعه داده اول در جدول۹ و نمودار دقت در شکل۳ و برای مجموعه داده دوم، نتایج در جدول۱۰ و نمودار دقت در شکل ۴ نشان‌داده‌شده است.

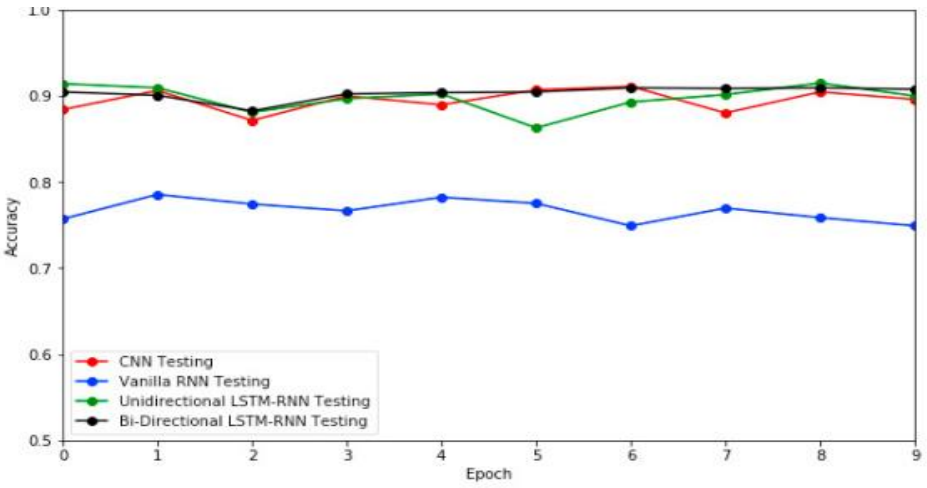
[55] Machine learning repository
[56] Train
[57] Validation
[58] Test
[59] Unique Token

جدول۹: نتایج ارزیابی برای مجموعه داده اول[۲۰]

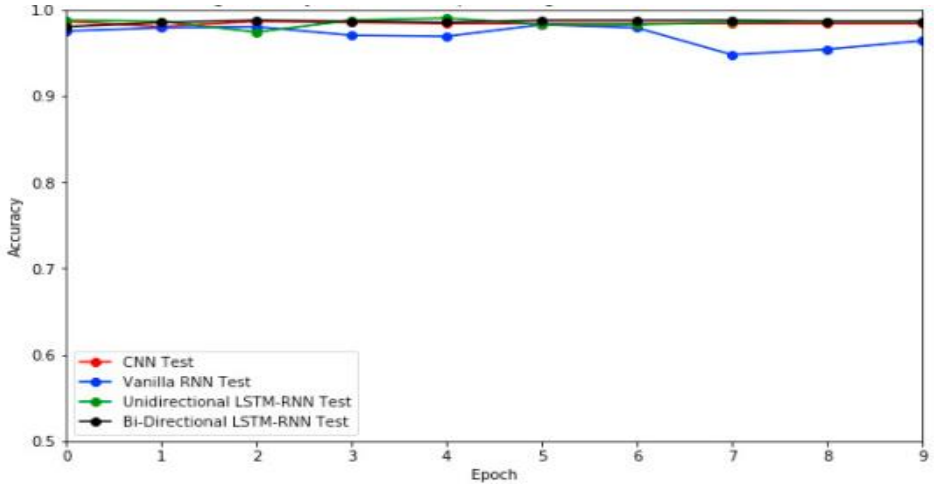
Model	Training Accuracy	Validation Accuracy	TestAccuracy
CNN	0.9942	0.8942	0.9077
Vanilla RNN	0.9971	0.7601	0.7822
Unidirectional LSTM-RNN	0.9939	0.8903	**0.9148**
Bi-directional LSTM-RNN	**0.9992**	**0.8974**	**0.9108**

جدول۱۰: نتایج ارزیابی برای مجموعه داده دوم[۲۰]

Model	Training Accuracy	Validation Accuracy	TestAccuracy
CNN	0.9992	0.9738	0.9833
Vanilla RNN	0.9971	0.9701	0.9638
Unidirectional LSTM-RNN	0.9996	0.9738	0.9863
Bi-directional LSTM-RNN	**1**	**0.9825**	**0.9875**



شکل۳: دقت شناسایی اخبار جعلی در مجموعه داده اول[۲۲]



شکل۴: دقت شناسایی اخبار جعلی در مجموعه داده دوم[۲۲]

جدول۱۱: نتایج ارزیابی مقاله[۲۴]

MODEL	PRECISION	RECALL	F_1
Logistic Regression	0.96	0.49	0.65
Feedforward Network	0.89	0.74	0.80
RNN (Vanilla)	0.91	0.56	0.70
GRUs	0.89	**0.79**	**0.84**
LSTMs	0.93	0.72	0.81
BiLSTMs	0.88	0.75	0.81
CNN with Max Pooling	0.87	0.44	0.58
CNN with Max Pooling and Attention	**0.97**	0.03	0.06

۱۳- نتیجه گیری

در تحقیق مورد نظر عملکرد الگوریتم ها و مدل های مختلف کلاسیک و پیشنهادی یادگیری ماشین و یادگیری شبکه های عصبی عمیق در مقالات مختلف را برر سی کردیم و به این نتیجه ر سیدیم که نتایج ارزیابی بر روی مدل های یادگیری شبکه های عصبی عمیق در اغلب اوقات موفق تر از مدل های کلاسیک یادگیری ماشین بوده اند.

۱۴- پیشنهادات

استفاده از روش های هیبریدی[۶۰] با ایده ترکیب الگوریتم های شبکه های عصبی عمیق با الگوریتم های کلاسیک یادگیری ماشین به عنوان الگوریتم های انسمبل[۶۱] می تواند در رفع نقاط ضعف هر کدام از این الگوریتم ها موثر و منجر به بهبودی در معیارهای مختلف سنجش از جمله دقت مدل شود.

[60] Hybrid

[61] Ensemble algorithms

مراجع

[1] Shrivastava, Sagar, Rishika Singh, Charu Jain, and Shivangi Kaushal. "A Research on Fake News Detection Using Machine Learning Algorithm." In *Smart Systems: Innovations in Computing*, pp. 273-287. Springer, Singapore, 2022.

[2] Ahmed, Sajjad, Knut Hinkelmann, and Flavio Corradini. "Combining machine learning with knowledge engineering to detect fake news in social networks-a survey." arXiv preprint arXiv:2201.08032 (2022).

[3] A. Taherkhani, G. Cosma, and T. M. McGinnity, "AdaBoost-CNN: An adaptive boosting algorithm for convolutional neural networks to classify multi-class imbalanced datasets using transfer learning," Neurocomputing, vol. 404, pp. 351-366, 2020

[4] P. Bahad, P. Saxena, and R. Kamal, "Fake News Detection using Bi-directional LSTM-Recurrent Neural Network," Procedia Computer Science, vol. 165, pp. 74-82, 2019.

[5] M. A. Wani and S. Jabin, "A sneak into the Devil's Colony-Fake Profiles in Online Social Networks," *arXiv preprint arXiv:1705.09929*, 2017.

[6] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD explorations newsletter*, vol. 19, no. 1, pp. 22-36, 2017.

[7] B. Wang, A. Zubiaga, M. Liakata, and R. J. a. p. a. Procter, "Making the most of tweet-inherent features for social spam detection on Twitter," 2015.

[8] C. Chen *et al.*, "Statistical features-based real-time detection of drifted twitter spam," vol. 12, no. 4, pp. 914-925.٢٠١٧ ,

[9] C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou, and G. Min, "Statistical features-based real-time detection of drifted twitter spam," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 4, pp. 914-925, 2016.

[10] S. Pouyanfar *et al.*, "A survey on deep learning: Algorithms, techniques, and applications," *ACM Computing Surveys (CSUR)*, vol. 51, no. 5, pp. 1-36, 2018.

[11] S. Sedhai and A. Sun, "Semi-supervised spam detection in Twitter stream," *IEEE Transactions on Computational Social Systems*, vol. 5, no. 1, pp. 169-175, 2017.

[12] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 2016, pp. 855-864 .

[13] S. K. Maity, S. KC, and A. Mukherjee, "Spam2vec: Learning biased embeddings for spam detection in twitter," in *Companion Proceedings of the The Web Conference 2018*, 2018, pp. 63-64 .

[14] W. Yin, K. Kann, M. Yu, and H. Schütze, "Comparative study of cnn and rnn for natural language processing," *arXiv preprint arXiv:1702.01923*, 2017

[15] T. T. Pham, "A study on deep learning for fake news detection," 2018.

[16] R. Oshikawa, J. Qian, and W. Y. Wang, "A survey on natural language processing for fake news detection," *arXiv preprint arXiv:1811.00770*, 2018.

[17] S. T. Allaparthi, G. Yaparla, and V. Pudi, "Sentiment and Semantic Deep Hierarchical Attention Neural Network for Fine Grained News Classification," in *2018 IEEE International Conference on Big Knowledge (ICBK)*, 2018: IEEE, pp. 65-72 .

[18] P. Zhou *et al.*, "Attention-based bidirectional long short-term memory networks for relation classification," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2016, pp. 207-212 .

[19] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119-139, 1997.

[20] 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 .

[21] R. E. Schapire, "Explaining adaboost," in *Empirical inference*: Springer, 2013, pp. 37-52.

[22] S. Bajaj, "The Pope Has a New Baby! Fake News Detection Using Deep Learning," ed: Tech. rep. Technical Report, Stanford Univ, 2017.

[23] T. Martensen "Uncovering Fake News with Machine Learning," ed: Blekinge Institute of Technology,Karlskrona, Sweden 2018.

[24] C. Boididou, S. Papadopoulos, M. Zampoglou, L. Apostolidis, O. Papadopoulou, and Y. Kompatsiaris, "Detection and visualization of misleading content on Twitter," *International Journal of Multimedia Information Retrieval*, vol. 7, no. 1, pp. 71-86, 2018.