Journal of Soft Computing and Information Technology (JSCIT)
Babol Noshirvani University of Technology, Babol, Iran
Journal Homepage: jscit.nit.ac.ir
Volume 10, Number 1, Spring 2021, pp. 13-23
Received: 06/14/2020, Revised: 09/29/2020, Accepted: 11/24/2020

Determination of Mental States from Texts Using Evolutionary Imperialist Competitive Algorithm and Convolution Neural Networks

Bahareh Golestanifar[1] and Abdolah Chalechale[2*]
1- Computer Dept., Eng. Faculty, Razi University, Kermanshah, Iran.
2*- Computer Dept., Eng. Faculty, Razi University, Kermanshah, Iran.
[1]golestany950@yahoo.com, and [2*] chalechale@razi.ac.ir

Corresponding author's address: Abdolah Chalechale, Faculty of Electrical and Computer Engineering, Razi University, Kermanshah, Iran, Post Code: 6714414971.

Abstract- The aim of this study is to investigate writings to find out the mood of people in typing texts. In this study, 14640 tweets related to airlines were used to analyze emotions in three categories: positive, negative and neutral. The novel proposed approach has three main steps. In the first step, we perform a pre-processing operation to purify the dataset. In the second step, using the Imperialist Competitive Algorithm (ICA), the main keywords from all the existing texts are extracted. Keywords are the words that have the most impact on categorization. Then, a convolution neural network (CNN) is exploited to extract more features. In the last step, classification, using a multilayer perceptron neural network (MLP), is applied. In the proposed new method, unlike the conventional methods in which words go to the next stage after preprocessing, we use the Imperialist Competitive Algorithm to extract the main words from all these words, which in turn causes There is a significant reduction in the volume of input words. Using this new proposed approach, we achieved precision, accuracy and recall of 0.990, 0.983 and 0.875, respectively. The experimental results indicated the superiority of the proposed method is comparison with other well-known approaches.

Keywords- Data Mining, Text Mining, Sentiment Analysis, Neural Networks, Convolutional Neural Network, Imperialist Competitive Algorithm.

مجله علمی پژوهشی رایانش نرم و فناوری اطلاعات
دانشگاه صنعتی نوشیروانی بابل
صفحه مجله: jscit.nit.ac.ir
جلد ۱۰، شماره ۱، بهار ۱۴۰۰ ، صفحه ۱۳-۲۳
دریافت: ۱۳۹۹/۰۳/۲۵، بازنگری: ۱۳۹۹/۰۷/۰۸، پذیرش: ۱۳۹۹/۰۹/۰۴

تحلیل و بررسی تعیین حالات روحی از روی متون با استفاده از الگوریتم تکاملی رقابت استعماری

بهاره گلستانی فر[۱]، عبدالله چاله چاله*[۲]
۱- دانشکده مهندسی برق و کامپیوتر، دانشگاه رازی، کرمانشاه، ایران.
*۲- دانشکده مهندسی برق و کامپیوتر، دانشگاه رازی، کرمانشاه، ایران.
[1]golestany950@yahoo.com, [2*] chalechale@razi.ac.ir

* نشانی نویسنده مسئول: عبدالله چاله چاله، کرمانشاه، خیابان دانشگاه، دانشگاه رازی، دانشکده مهندسی کامپیوتر، کد پستی ۶۷۱۴۴۱۴۹۷۱.

چکیده- هدف از پژوهش حاضر متن‌کاوی جهت پی بردن به حالت روحی افراد در تایپ متون است. در این پژوهش از ۱۴۶۴۰ توئیت در رابطه با خطوط هوایی جهت تحلیل احساسات در سه دسته‌ی مثبت، منفی و خنثی استفاده شده است. طرح پیشنهادی جدید دارای سه مرحله اصلی است. در مرحله اول پیش پردازشی به منظور آماده سازی پایگاه داده انجام می‌گیرد. در مرحله دوم با استفاده از الگوریتم رقابت استعماری کلمات اصلی را از بین تمامی کلمات موجود استخراج می‌شود. منظور از کلمات اصلی کلماتی هستند که بیشترین تأثیر را برای دسته‌بندی دارند. سپس با استفاده از یک شبکه‌ی عصبی پیچشی اقدام به استخراج ویژگی‌های مناسب صورت می‌گیرد. سپس درمرحله آخر با استفاده از یک شبکه‌ی عصبی پرسپترون چندلایه (MLP) طبقه بندی انجام می‌گیرد. در روش پیشنهادی جدید، برخلاف روش‌های مرسوم که کلمات بعد از پیش پردازش به مرحله‌ی بعد راه پیدا می‌کنند، با استفاده از الگوریتم رقابت استعماری از میان تمامی این کلمات ،کلمات اصلی را استخراج می‌کنیم که این کار خود سبب کاهش قابل توجهی از حجم کلمات ورودی می‌شود. طرح پیشنهادی جدید در آزمایش تجربی توانست به پارامترهای دقت ۰/۹۹۰، صحت ۰/۹۸۳ و فراخوانی ۰/۸۷۵ برسد. این نتایج نشان‌دهنده‌ی برتری طرح پیشنهادی این مقاله در مقایسه با سایر روش های قبلی می‌باشد.

واژه‌های کلیدی: داده‌کاوی، متن‌کاوی، تحلیل احساسات، شبکه‌های عصبی، شبکه عصبی پیچشی، الگوریتم رقابت استعماری.

۱- مقدمه

هدف اصلی انسان از جمع‌آوری اطلاعات را می‌توان فهمیدن تفکر سایر انسان‌ها دانست. این تمایل ناخودآگاه سبب کشش پژوهشگران به سمت تحلیل اطلاعات جهت درک و تجزیه ذهن سایر انسان‌ها، شده است. امروزه با توجه به پیشرفت‌ها و فراهم آوردن بسترهای اطلاعاتی مانند اینترنت، شبکه‌های اجتماعی و غیره می‌توان اطلاعات موردنیاز خود را به راحتی جمع‌آوری کرد. تشخیص احساسات از این اطلاعات یکی از جنبه‌های مهم تحلیل اطلاعات محسوب می‌شود و تا این لحظه بسیاری از محققین به این سمت‌وسو هدایت شده‌اند. احساسات ناشی از ترکیبی پیچیده از هورمون‌ها و ناخودآگاه ذهن انسان است که تنها با تلاش‌های بسیار می‌توان آن را کنترل کرد بنابراین می‌توان احساسات را به‌راحتی از اطلاعات به دست آورد.

تشخیص احساس را می‌توان در قالب‌های متفاوتی مانند گفتار، چهره و متن انجام داد که در این پژوهش تمرکز بر روی متن قرار

دارد زیرا بسیاری از ابزارهایی که انسان‌ها با یکدیگر ارتباط برقرار می‌کنند از طریق نوشتار یا متن است.

یکی از منابع اطلاعاتی مهم جهت جمع‌آوری اطلاعات شبکه‌های اجتماعی هستند. طبق آخرین آمارها [۱] شبکه‌ی اجتماعی فیس‌بوک حدوداً دارای۱ میلیارد کاربر فعال بوده که حداقل یک‌بار در ماه وارد صفحه شخصی خود می‌شوند. این کاربران در یک دقیقه حدود ۴۱۰۰۰ بار متنی را در فیس‌بوک می‌نویسند. در شبکه اجتماعی توئیتر روزانه ۵۰۰ میلیون توئیت به ۳۳ زبان زنده دنیا توسط کاربران نوشته می‌شود. در میان شبکه‌های اجتماعی موجود توئیتر از اهمیت بیشتری برخوردار است. یکی از دلایل آن محدودیت تعداد کاراکترهای استفاده شده در هر توئیت است. این محدودیت باعث می‌شود که کاربران منظور و مفهوم خود را در جملات کوتاه‌تر و با استفاده از کلمات کلیدی با ارزش‌تر بیان نمایند.

برخی محققین متن‌کاوی را به‌عنوان ابزاری برای جستجو کردن اطلاعات "باکیفیت" در حجم زیادی از داده‌ها تعریف می‌کنند. در متن‌کاوی"باکیفیت" به مجموعه‌ای از ارتباط داشتن، نوپدید بودن و موردتوجه بودن اشاره می‌شود. متن‌کاوی به‌عنوان آنالیز متفکرانه متن یا داده‌کاوی متن نیز بیان می‌شود. برای انجام فرایند متن‌کاوی زمینه‌های گوناگون تحقیقی وجود دارد. در حال حاضر، تعیین مناسب‌ترین روش جهت به‌دست آوردن حالت بهینه یک چالش بزرگ محسوب می‌شود. اگر فقط به داده در سطح وب، مانند شبکه‌های اجتماعی نگاه کنیم، می‌توان متوجه شد که چشم‌انداز داده‌ها به‌صورت غیر ساخت‌یافته است. این داده‌ها برای شرکت‌ها، دولت‌ها، خدمات مالی، حوزه‌ی توسعه تجارت [۲]، سازمان‌های دفاع و پژوهشگران علمی بسیار ارزشمند است.

در این مقاله به ارائه‌ی روشی جدید جهت تعیین حالات روحی از روی متون پرداخته شده است. در روش پیشنهادی در ابتدا با استفاده از الگوریتم فراابتکاری رقابت استعماری کلمات اصلی از بین تمامی کلمات موجود در متن استخراج شده است. سپس با استفاده از شبکه عصبی پیچشی اقدام به یادگیری ویژگی شده و در انتها با استفاده از شبکه عصبی پرسپترون چندلایه اقدام به دسته بندی می‌شود. با استفاده از الگوریتم رقابت استعماری و نحوه‌ی تعیین لغات اصلی از سایر لغات و شبکه‌ی عصبی پیچشی حجم داده‌های ورودی جهت طبقه‌بندی کاهش پیدا کرده است و همچنین نتایج نیز بهبود یافته است.

در ادامه مقاله در بخش دوم کارهای مرتبط بیان می‌شود. در بخش سوم فرآیند متن‌کاوی شرح داده می‌شود. در بخش چهارم تکنیک‌های موجود در آنالیز احساسات بیان می‌شود سپس در بخش پنجم الگوریتم رقابت استعماری تشریح می‌شود. در بخش ششم به‌تشریح روش پیشنهادی اولیه (IM[1]) و روش پیشنهادی نهایی بهبود یافته (ICM[2]) می‌پردازیم. در بخش هفتم معیارهای ارزیابی بیان می‌شوند. در بخش بعد به ارزیابی نتایج می‌پردازیم. در بخش آخر هم نتیجه‌گیری و کارهای آینده بیان می‌شود.

۲- کارهای مرتبط

در این بخش به بررسی کارهایی که پیشتر انجام شده است خواهیم پرداخت. هر چند تعداد آثار موجود در زمینه تحلیل احساس از متن، فراوان است ولی در این بخش به تعدادی از تحقیقات اشاره خواهد شد.

در [۳] با به‌کارگیری انواع تکنیک‌های یادگیری ماشین در بستر توئیتر به این نتیجه رسیده‌اند که روش‌های به کار گرفته‌شده برای بردار ویژگی جدید دارای صحت[3] مشابه هستند اما طبقه بند بیز ساده[4] دارای دقت[5] بهتری در مقایسه با سایر دسته‌بندها است. در [۴] جهت حل مشکل فهم ویژگی‌های معنایی و بهبود دقت دسته‌بندی، یک متد LSTM[6] بهبود داده‌شده به نام AC-LSTM[7] ارائه شده است. محققین با پیاده‌سازی مدل پیشنهادی بر روی هفت پایگاه داده به این نتیجه رسیده‌اند که مدل AC-LSTM معنا را به‌درستی درک و کارایی متد LSTM را هم بهبود می‌دهد.

در [۵] سعی شده برتری روش تولید یک لغت‌نامه احساسی در حوزه‌ی موردنظر (DSELs[8]) را جهت بالا بردن کیفیت تخمین محدوده‌ی احساسی کلمات، نسبت به سایر روش‌ها که از لغت‌نامه‌های آماده استفاده می‌کنند (GPELs[9]) نشان دهد. همچنین با بررسی‌های عمیق‌تر، مشخص گردیده است که روش DSEL در تولید موضوع‌هایی با ابهام کمتر، نسبت به روش‌های GPEL موفق‌تر بوده است.

در [۶،۷] باهدف آنالیز احساسات کاربران، بر روی متون فارسی، تحقیق شده است. تکنیک‌های ارائه‌شده به ترتیب با استفاده از روش‌های مبتنی بر شباهت معنایی و تکنیک‌های ریشه‌یابی بر روی داده‌های چند کلاسه، استفاده از شبکه‌ی LSTM و استخراج ویژگی‌های ترکیبی بر روی داده‌های دو کلاسه و چند کلاسه توانسته‌اند به‌ترتیب به دقت‌های ۸۸٪ و ۸۳٪ در حالت دو کلاسه و ۸۰٪ در حالت چند کلاسه دست پیدا کنند.

در جدول ۱ کارهای مرتبط بر اساس تکنیک‌های تعیین حالات روحی دسته‌بندی و نمایش داده شده است.

جدول۱: مقایسه کارهای مرتبط.

نویسندگان	پایگاه داده	تکنیک تعیین حالات روحی	سال
Neethu M S, Rajasree R [3]	توئیتر	یادگیری ماشین	۲۰۱۳
Gang Liu, Jiabao Guo [4]	مجموعه‌ای از داده‌های ذهنی	یادگیری عمیق(-AC BiLSTM)	۲۰۱۹
Anil Bandhakavi, Nirmalie Wiratunga Lamaute & et al [5]	توئیتر	بر اساس لغت نامه	۲۰۱۷
سارا منتظریان، سید محمد حسین معطر [۶]	پایگاه داده‌ی آزمایشگاه دانشگاه امیرکبیر	روش مبتنی بر شباهت معنایی	۲۰۱۵
Shima Zobeidi, Marjan Naderan & et al [7]	دیجی کالا	یادگیری عمیق	۲۰۱۹

۳- فرآیند متن‌کاوی

به‌طورکلی می‌توان فرآیند متن‌کاوی را به دو مرحله‌ی کلی تقسیم کرد: پیش پردازش و استخراج دانش. پیش‌پردازش داده‌ها گامی مهم در فرآیند متن‌کاوی است که می‌تواند سرعت و دقت کار را افزایش دهد. دلیل انجام این گام این است که داده‌ها همیشه از کیفیت لازم برخوردار نیستند و تمام کلمات موجود در متن نیز مفید واقع نمی‌شوند [۹,۸]. روش‌هایی که در این زمینه وجود دارد شامل: پاک‌سازی داده‌ها، تجمیع داده‌ها، کاهش داده‌ها و تبدیل داده‌ها است.

۴- تکنیک‌های آنالیز احساسات

کشف احساسات از متن در محدوده‌های متنوعی ازجمله نظرات در شبکه‌های اجتماعی، عناوین اخبار، محتوای وبلاگ‌ها، اسناد، یادداشت‌ها و ایمیل‌ها به‌کار برده می‌شود. همچنین این تحلیل می‌تواند در سطح کلمه، عبارت، جمله یا سند انجام شود.

در [۱۱،۱۲] تکنیک‌هایی که در زمینه‌ی آنالیز احساسات وجود دارد را به سه دسته‌ی کلی بر اساس لغت‌نامه، یادگیری ماشین و روش‌های ترکیبی تقسیم کرده‌اند که در ادامه به شرح مختصری از هریک پرداخته می‌شود.

۴-۱- تکنیک‌های مبتنی بر لغت‌نامه

توجه به دو نکته‌ی تشخیص میل (جهت) کلمه و جمع‌آوری میل کلمات برای تعریف تمایل کل متن بسیار مهم است. جهت ایجاد یک لغت‌نامه در یک محدوده‌ی خاص می‌توان در ابتدا کلمات آن گستره را به‌صورت محدود جمع‌آوری کرد و سپس با استفاده از کلمات هم‌معنی و متقابل در سایر لغت‌نامه‌ها این واژه‌نامه را گسترش داد. در انتها می‌توان از تکنیک‌های بیشترین نمرات و حاصل جمع نمرات مثبت و منفی جهت جمع‌کردن نمرات کل استفاده کرد. در اینجا نمرات، تاثیر کمی جهت لغات را نشان می‌دهد. لغت‌نامه‌ها می‌توانند به‌صورت دستی توسط خود نویسنده، ایجاد شوند و یا از لغت‌نامه‌های آماده‌ای مانند Crowdsourcing [۱۳] و Wordnet [۱۴] استفاده شود.

تکنیک‌های مبتنی بر لغت‌نامه دارای دو مشکل عمده هستند:

۱. ابهام در ترکیب زبان، دلیل اصلی عدم استفاده از این تکنیک است. به‌عنوان نمونه می‌توان به جملاتی اشاره کرد که باوجود داشتن کلمات منفی، گرایشی مثبت دارند [۱۵].
۲. تشبیه، کنایه و طعنه سد دیگری است که در هنگام کار با این تکنیک با آن مواجه هستیم. البته واژه‌نامه‌هایی که در این گستره هستند می‌تواند بسیار کمک کند اما در مورد تشبیه و طعنه این را نمی‌توان گفت. N-gram روشی است که در این حالت می‌توان از آن استفاده کرد.

جمله‌ی زیر را در نظر بگیرید:

" فکر می‌کردم موجودات یا انسان هستند یا نیستند. نمی‌دانستم یکی می‌تواند مقداری انسان باشد"

این جمله نمونه‌ی خوبی از پیچیدگی ساختار زبان است.

۴-۲- روش‌های یادگیری ماشین

در این‌گونه تکنیک‌ها نیازی به در نظر گرفتن ساختار گرامری جمله نیست. در این رویکرد از الگوریتم‌های مشهور یادگیری و مبتنی بر ویژگی‌های زبانی استفاده می‌شود. بنابراین باید از یک مجموعه داده‌ی آموزشی مناسب جهت یادگیری استفاده کرد. درنتیجه می‌توان پی برد که مهم‌ترین قسمت‌های این تکنیک، انتخاب و استخراج ویژگی است. این تکنیک‌ها خود به دو دسته‌ی با نظارت [۱۰] و بدون نظارت [۱۱] تقسیم می‌شوند.

شبکه‌های عصبی پیچشی[۱۲](CNN) نمونه‌ای از شبکه‌های عصبی عمیق است که در یادگیری ماشین استفاده می‌شود. یادگیری عمیق زیرمجموعه یادگیری ماشین است. این شبکه به‌دلیل وجود لایه‌های بیشتر (لایه‌های بیشتر نشان‌دهنده‌ی عمق آن شبکه است) می‌تواند ویژگی‌های بیشتری را با دقت بالاتری استخراج و از آن‌ها جهت عملیات‌های متفاوتی مانند دسته‌بندی استفاده کند. یکی از دلایل استفاده از این شبکه‌ها در بسیاری از کارها به حداقل رساندن عملیات پیش‌پردازش است. یعنی این شبکه می‌تواند ویژگی‌ها و الگوهایی را به دست آورد که در شبکه‌های عصبی دیگر

به‌راحتی به دست نمی‌آیند [۱۶]. معمولاً این شبکه‌ها را با نام شبکه‌های عصبی پیچشی یا کانولوشن به رسمیت می‌شناسند اما گاهی این شبکه‌ها را با نام تغییرناپذیر با انتقال[۱۳] یا تغییرناپذیر با فضا [۱۴] نیز می‌شناسند [۱۷،۱۸].

۴-۳- روش‌های ترکیبی

در برخی از پژوهش‌ها از روش ترکیبی استفاده شده است. بدین ترتیب که داده‌هایی که در ابتدا با تکیه بر روش مبتنی بر لغت‌نامه بار احساسی آن‌ها تعیین شده، به‌عنوان داده‌های ورودی برای آموزش مدل در الگوریتم‌های یادگیری ماشین استفاده می‌شود. به‌عنوان مثال در [۱۹] رویکردی ترکیبی برای تعیین شدت نظرات (مثبت یا منفی) نسبت به هر کتاب پیاده‌سازی شده است. مهم‌ترین مزیت این روش بازده بالا در دقت نتایج حاصل از ترکیب دسته‌بندهای آموزش‌دیده و اطلاعات زبان‌شناسی غنی حاصل از لغت‌نامه‌ها است.

۵- الگوریتم رقابت استعماری

الگوریتم رقابت استعماری [۱۵] ازجمله‌ی الگوریتم‌های بهینه‌سازی است که از پدیده‌ی اجتماعی- سیاسی گرفته‌شده است. مدت زیادی از معرفی این الگوریتم نمی‌گذرد اما در بسیاری از زمینه‌های صنعتی، مکانیکی، طراحی، کنترل و کشف مسیرهای بهینه استفاده‌شده است [۲۰]. از لحاظ کارایی و نتیجه هیچ روشی لزوماً بهتر از روش‌های دیگر نیست. مهم‌ترین مزیت این الگوریتم این است که، نسبت به الگوریتم‌های شناخته شده، جدیدتر و نسبت به الگوریتم‌های جدید، شناخته شده‌تر است. الگوریتم رقابت استعماری نه آنقدر قدیمی است که در توسعه و ایجاد تغییرات در آن با مشکل مواجه شویم و نه آنقدر جدید است که در شناساندن و معرفی آن و همچنین یافتن مستندات آموزشی آن به مشکل برخورد کنیم. از طرفی در مقایسه با الگوریتم‌های جدیدتر مانند الگوریتم خفاش [۱۶] که با کاهش دقت روبرو می‌شود و ممکن است دچار رکود پس از چند مرحله اول شود، بهتر عمل کرده است. همچنین با بررسی تحقیق [۲۱] می‌توان نتیجه گیری کرد که در سال‌های اخیر نیز این الگوریتم هنوز در بین سایر الگوریتم‌ها، توانایی خود را حفظ کرده است.

این الگوریتم به این نحو عمل می‌کند که در ابتدا یکسری جواب‌های کاندید به‌صورت تصادفی ایجاد می‌شوند. این‌ها درواقع راه‌حل‌های اولیه یا کاندیدهای اولیه‌ای هستند که دقیقاً معادل کروموزوم‌ها در الگوریتم ژنتیک می‌باشند. در این الگوریتم راه‌حل‌های ابتدایی " کشور" نامیده می‌شوند.

گام بعدی انتخاب بهترین‌ها از بین کاندیدها (نخبه‌ها در ژنتیک) است. به‌عنوان مثال می‌توان ده کاندید اول را به‌عنوان امپریالیست[۱۷] در نظر گرفت سپس بقیه کاندیدها را بر طبق یک پروسه بین این امپریالیست‌ها به‌عنوان مستعمره[۱۸] پخش کرد. سپس عملیات سیاست جذب اجرا می‌شود. طی این سیاست مستعمرات به سمت امپریالیست‌ها حرکت می‌کنند. عملگر انقلاب هم بعد از این پروسه انجام می‌شود. در این پروسه مستعمرات به‌صورت تصادفی جایشان را عوض می‌کنند و اگر مستعمره به‌جای بهتری از امپریالیست برسد کل امپراتوری را تصاحب می‌کند.

سپس هزینه کل امپراتوری از طریق رابطه ۱ محاسبه می‌شود.

$$T.C._n = cost(imperialist_n) + \xi\, mean\{cost\ (colonies\ of\ empire_n)\} \qquad (۱)$$

هزینه کل یک امپراتوری ($T.C._n$) برابر است باقدرت کشور استعمارگر ($cost(imperialist_n)$) به‌علاوه میانگین قدرت مستعمره‌های امپراتوری ($colonies\ of\ empire_n$) ضرب در پارامتر ξ که عددی بسیار کوچک است. بخش بعد رقابت امپریالیستی است که در آن یکی از مستعمرات ضعیف‌ترین امپراتوری در جریان رقابت قرار می‌گیرد و بقیه جهت به دست آوردن آن رقابت می‌کنند. جهت محاسبه‌ی این رقابت رابطه ۲ در نظر گرفته شده است.

$$N.T.C._n = max\{T.C._i\} - T.C._n \qquad (۲)$$

در این رابطه $T.C._n$ هزینه کل امپراتوری nام و $N.T.C._n$ هم هزینه کل نرمالیزه شده آن امپراتوری است. هر امپراتوری که $T.C._n$ کمتری داشته باشد $N.T.C._n$ بیشتری خواهد داشت. امپراتوری با کمترین هزینه، دارای بیشترین قدرت است. سپس احتمال تصاحب هر امپراتوری را می‌توان از طریق رابطه ۳ محاسبه کرد.

$$P_{p_n} = \frac{N.T.C._n}{\sum_{i=1}^{N_{imp}} N.T.C._i} \qquad (۳)$$

حال که احتمال تصاحب برای هر امپراتوری به دست آمده است نیازمند روشی هستیم تا بتوان مستعمرات را با احتمال به‌دست‌آمده به استعمارگران داد. در ابتدا بردار P مطابق رابطه ۴ جهت اختصاص مستعمرات با توجه به احتمال تصاحب هر امپراتوری، تشکیل شده است.

$$P = [p_{p_1}, p_{p_2}, p_{p_3}, \dots, p_{p_{N_{imp}}}] \qquad (۴)$$

این بردار شامل احتمال تصاحب امپراتوری‌ها است. در مرحله بعد بردار R به‌صورت تصادفی هم سایز با بردار P مطابق رابطه ۵ تشکیل شده است. آرایه‌های این بردار، اعدادی تصادفی در بازه [۰و۱] هستند.

$$R = [r_1, r_2, r_3, \dots, r_{N_{imp}}] \qquad (۵)$$

$$R = r_1, r_2, r_3, \dots, r_{N_{imp}} \sim U(0,1)$$

سپس بردار D طبق رابطه ۶ ایجاد می‌شود.

$$D = P - R = \left[D_1, D_2, D_3, \dots, D_{N_{imp}}\right] = \qquad (۶)$$

$$\left[p_{p_1} - r_1, p_{p_2} - r_2, p_{p_3} - r_3, \dots, p_{p_{N_{imp}}} - r_{N_{imp}}\right]$$

حال با استفاده از این بردار مستعمرات به امپراتوری داده می‌شود که اندیس مربوط به آن در بردار D بزرگ‌تر از بقیه باشد. امپراتوری که بیشترین احتمال تصاحب را داشته باشد، با احتمال بیشتری اندیس مربوط به آن در بردار D، بیشترین مقدار را خواهد داشت.

اگر امپریالیستی در کل این جریان تمام مستعمرات خودش را از دست بدهد عملاً دیگر معنایی ندارد و از بین می‌رود. البته می‌توان یکسری شرایط توقف هم داشت که در صورت برقراری، کار تمام می‌شود.

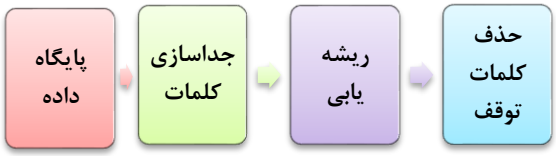
۶- روش IM و ICM

در روش IM ابتدا پایگاه داده‌ی موردنظر جهت دسته‌بندی بدون اعمال شبکه‌ی عصبی پیچشی دسته‌بندی می‌شود. هدف از این گام نمایش تأثیر شبکه عصبی پیچشی بر روی پایگاه داده‌ی موردنظر است. در انتها روش پیشنهادی نهایی بهبود یافته با روش پیشنهادی اولیه (دسته‌بندی احساسات بدون اعمال شبکه عصبی پیچشی) مقایسه می‌شود.

روش ICM برای تشخیص احساس از روی متن شامل سه فاز مختلف است. در فاز اول ابتدا پیش‌پردازش متن انجام می‌شود. عملیات پیش‌پردازش در راستای بهبود بخشیدن به استخراج اطلاعات است. در فاز دوم، کلمات باقیمانده از پیش‌پردازش به‌عنوان ورودی به روش بهینه‌سازی رقابت استعماری داده می‌شود و در فاز سوم عملیات دسته‌بندی انجام می‌شود. در ادامه هر فاز به‌صورت کامل شرح داده شده است.

۶-۱- پیش‌پردازش

در این مرحله سه عملیات اصلی به‌وسیله‌ی نرم‌افزار پایتون (IDLE[۱۹]) بر روی متون انجام می‌شود. مراحل انجام‌شده بر روی



شکل ۱: مراحل پیش‌پردازش.

پایگاه داده در شکل ۱ نمایش داده‌شده است.

- جداسازی کلمات

در این مرحله متن برحسب نقاط و فاصله‌های بین کلمات، به واحدهای کوچک‌تر به نام توکن تقسیم می‌شود. توکن‌ها معمولاً شامل کلمات یا عبارت‌های چندکلمه‌ای می‌شوند. جهت انجام جدا سازی کلمات[۲۰] با توجه به ابزارهای معرفی‌شده در مقاله [۲۲] از NLTK Word Tokenize استفاده شده است.

- ریشه‌یابی

در این مرحله کلمات به ریشه‌ی اصلی آن‌ها نگاشت داده می‌شوند. این کار سبب افزایش سرعت کار و بالا رفتن دقت می‌شود. یکی دیگر از مزایای ریشه‌یابی[۲۱] کاهش ابعاد است [۲۳]. البته NLTK ریشه‌یاب‌های زیادی ازجمله Porter Stemmer، Lancaster Stemmer و Snowball Stemmer را شامل می‌شود که در این پژوهش از Porter استفاده شده است.

- حذف کلمات توقف

جملات همیشه حاوی یکسری علائم نگارشی مانند علامت سؤال، علامت تعجب و ویرگول هستند که ارزش معنایی خاصی ندارند و سبب کاهش سرعت می‌شوند.

پس یکی از راه‌های افزایش سرعت، حذف کلمات رایج و بدون معنای خاص یا همان کلمات توقف [۲۲] است.

۶-۲- استخراج کلمات اصلی

در این مرحله کلمات باقیمانده از پیش‌پردازش به‌عنوان ورودی به روش بهینه‌سازی رقابت استعماری داده می‌شود. روش رقابت استعماری طی تکرارهای مختلف اقدام به ایجاد راه‌حل‌های جدیدی برای مسئله بهینه‌سازی می‌نماید که باید توسط یک تابع

برازش، میزان مناسب بودن این راه‌حل تعیین شود تا در طول زمان راه‌حل‌های بهتر با برازش مطلوب‌تر جایگزین راه‌حل‌های نامناسب قبلی شود و درنهایت به جواب نزدیک به بهینه دست یافت. در اینجا هر راه‌حل مسئله به‌صورت دو بردار تحت عناوین Solution و Been است که مقادیر بردار Solution شماره اندیس لغات در مجموعه کلمات اصلی استخراج‌شده از مجموع توئیت‌ها است. مقادیر بردار Been نیز صفر و یک است تا در ترکیب با بردار Solution به ایجاد راه‌حل‌هایی با طول‌های مختلف منجر شود. نمونه‌ای از این راه‌حل‌ها در جدول ۲ نشان داده شده است.

جدول۲: شکل راه‌حل درروش فراابتکاری رقابت استعماری.

Solution	1	4	2	7	6	3	5
Been	1	0	1	1	1	0	0

درواقع این راه‌حل، معادل راه‌حل در جدول ۳ است.

جدول۳ : معادل راه‌حل درروش فراابتکاری رقابت استعماری.

Solution	1	2	7	6

سپس، ضریب همبستگی پیرسون مطابق رابطه ۷ به‌عنوان میزان همبستگی کلمه با برچسب نمونه (احساس فرد) محاسبه می‌شود. جهت نمایش نوع رابطه‌ی یک متغیر کمی با متغیر کمی دیگر می‌توان از این ضریب استفاده کرد. صفر بودن نشان‌دهنده‌ی عدم وجود رابطه و مقادیر بین ۱ و ۱- نشان‌دهنده‌ی شدت رابطه است.

$$corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \qquad (۷)$$

در رابطه ۷، E عملگر امید ریاضی، cov به معنای کوواریانس، $corr$ نماد معمول برای همبستگی پیرسون و σ نماد انحراف معیار است. درنهایت به ازای هر کلمه یک مقدار عددی به‌عنوان میزان همبستگی ایجاد می‌شود که در جدول ۴ نشان داده شده است.

جدول۴:میزان همبستگی کلمات.

Good	Nice	Bad	Service	Manage	...
0.5	0.8	0.7	0.1	0.2	...

در گام بعد با استفاده از روش بهینه‌سازی رقابت استعماری، زیرمجموعه‌ای از کلمات تأثیرگذار در تشخیص احساسات تعیین می‌شود. برای تعیین برازش راه‌حل‌های ارائه‌شده از رابطه ۸ استفاده خواهد شد که میزان مطلوب بودن هر راه‌حل را بر اساس مجموع همبستگی کلمات موجود در راه حل با برچسب مربوطه و همچنین طول راه‌حل تعیین می‌نماید. در اینجا هر چه مجموع مقادیر همبستگی کلمات بیشتر و طول راه‌حل (تعداد کلمات) کمتر باشد آنگاه مقدار برازش راه‌حل عدد بزرگ‌تری خواهد بود. این تابع برازش برای روش رقابت استعماری به‌صورت رابطه ۸ است.

$$fit(sol_i) = \frac{\sum_{w \in sol_i} cov(w,L(w))}{|sol_i|} \qquad (۸)$$

به‌طوری‌که $L(w)$ برچسب کلمه w است و $cov(w,L(w))$ میزان همبستگی پیرسون بین کلمه w (متعلق به راه‌حل sol_i) و برچسب این نمونه است.

پس از تعیین این لغات، توئیت‌ها بر اساس لغات تأثیرگذار به‌دست‌آمده از الگوریتم رقابت استعماری دوباره‌نویسی می‌شوند.

شکل ۲ بازنویسی دوباره یک توئیت را نمایش می‌دهد.

```
>> tweet(:,:,1)

     0     0     1     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0     0     0     0     0
     1     1     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0     0     0     0     0
```

شکل ۲: نمایش یک توئیت با استفاده از کلمات برگزیده.

حال، توئیت‌ها به ماتریس‌های مربعی دوبعدی مانند تصاویر تبدیل شده‌اند که هریک، اندازه یکسانی دارند تا برای یادگیری ویژگی به‌عنوان ورودی به شبکه عصبی پیچشی داده شوند. به عبارت دیگر، هدف ایجاد یک آرایه دوبعدی نرمال‌سازی شده به‌صورت جدول ۵ از توئیت‌های مختلف است که عناصر آن مقادیر شماره ترتیب وقوع یک کلمه در متن توئیت با استفاده از مقادیر صفر و یک باشد (با این کار پیچیدگی محاسباتی بسیار کاهش می‌یابد زیرا محاسبات محدود به مقادیر باینری می‌شوند).

جدول ۵: تبدیل دنباله کلمات اصلی توئیت‌های کاربران به یک ماتریس دوبعدی برای استفاده در CNN.

	1	2	3	4	5	6	7	8	9	10
good	1	0	0	0	0	0	0	0	0	0
nice	0	0	0	0	0	0	0	0	0	0
bad	0	0	0	0	0	0	0	0	0	0
service	0	1	0	0	0	0	0	0	0	0
manage	0	0	0	0	0	0	0	0	0	0
delay	0	0	0	1	0	0	0	0	0	0
noise	0	0	0	0	0	0	0	0	0	0
very	0	0	0	0	0	0	0	0	0	0
without	0	0	1	0	0	0	0	0	0	0
...	0	0	0	0	0	0	0	0	0	0
Outputs	Good, service, without, delay									

در اینجا اندازه مجموعه بهینه استخراج شده از روش رقابت استعماری برابر تعداد سطرها و برابر حداکثر تعداد کلمات اصلی در طولانی‌ترین توئیت موجود در داده‌ی آموزشی است. با توجه به این‌که این ماتریس می‌تواند یک ماتریس اسپارس باشد، می‌توان از عملگرهای اسپارس برای کاهش هرچه بیشتر پیچیدگی محاسباتی و زمان اجرای تشخیص استفاده کرد.

پس از این مرحله تمام توئیت‌ها به مثابه تصاویری تبدیل شده‌اند که هر یک اندازه یکسانی دارند با این تفاوت که در تصاویر، مقادیر در بازه صفر تا ۲۵۵ و باوجود سه کانال رنگ (در تصاویر RGB) است اما در اینجا یک آرایه دوبعدی با مقادیر دودویی ایجاد خواهد شد. حال می‌توان از شبکه عصبی پیچشی برای یادگیری ویژگی استفاده کرد.

در گام بعد هر توئیت با استفاده از ۱۰ فیلتر مختلف به ده ماتریس جدید تبدیل می‌گردد تا یادگیری ویژگی برای افزایش دقت روش با استفاده از تعداد بیشتری از مقادیر انجام شود. در شکل ۳ حاصل اعمال فیلترهای مختلف بر روی ماتریس توئیت در لایه کانولوشن نشان داده‌شده است. پارامترهای استفاده شده در الگوریتم رقابت

شکل۳: کانوالو نمودن نمایش ماتریسی یک توئیت در فیلترهای مختلف.

استعماری در جدول ۶ نمایش داده شده است.

جدول۶: پارامترهای الگوریتم رقابت استعماری.

ICA		
npop=50	تعداد کشورها	
nemp=10	تعداد امپراطوری‌ها	
MaxIt=30	تعداد تکرار	
Zeta=./1	ضریب میانگین قدرت کلونی‌های یک استعمارگر	

شبکه عصبی پیچشی در نظر گرفته‌شده برای فاز سوم و یادگیری شبکه عصبی پیچشی، شامل دو لایه پیچشی[۲۳]، دو لایه تجمعی[24] و درنهایت یک لایه کاملاً همبند[۲۵] است که لایه‌های پیچشی و تجمعی به‌صورت در هم‌گذاری شده (تکرار دو سطح از ترکیب یک‌لایه پیچشی و یک‌لایه تجمعی) استفاده‌شده است.

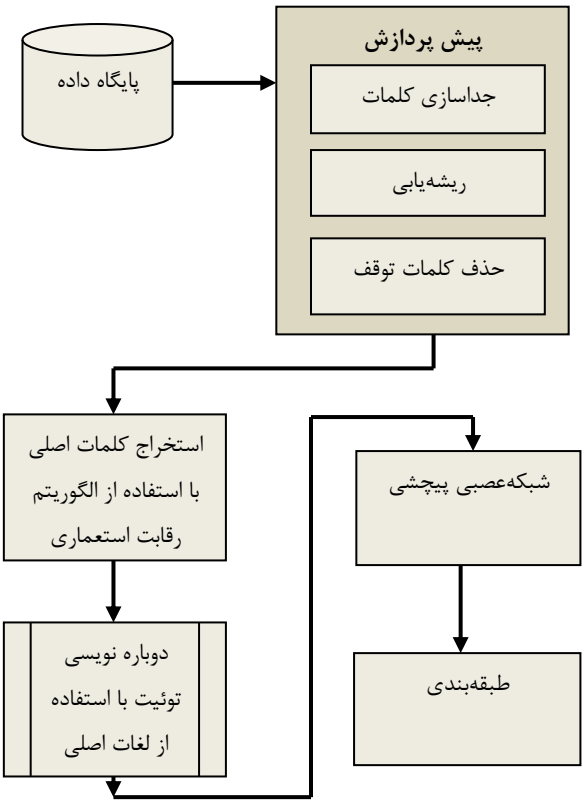
لایه‌های کاملاً همبند، لایه‌های طبقه‌بندی نهایی هستند که ویژگی‌های خروجی به‌صورت ضرب بردار ماتریس وزن کاملاً متصل و بردار ورودی ویژگی محاسبه می‌شوند. به دنبال اغلب لایه‌های پیچشی و کاملاً همبند، توابع فعال‌سازی غیرخطی مانند تانژانت هیپربولیک، سیگموئید یا واحد یکسو شده خطی (RELU[۲۶]) آورده می‌شوند. RELU، با تعریف $y = \max(x, 0)$، به دلیل همگرایی سریع و نیز پیاده‌سازی سخت‌افزاری جامع، انتخاب رایجی برای تابع فعال‌سازی است.

۶-۳- دسته‌بندی

در این مرحله ویژگی‌های استخراج‌شده در لایه کاملاً همبند به‌عنوان ورودی به یک دسته‌بند داده می‌شوند تا دقت روش مشخص شود. برای این کار از شبکه عصبی پرسپترون چندلایه (MLP[27]) استفاده شده است. شبکه عصبی پرسپترون با استفاده از توابع آموزشی مختلف مورد بررسی قرار گرفته است. تابع بیزین[۲۸] به زمان بیشتری احتیاج دارد. تابع شیب توام مقیاس شده [۲۹]نسبت به سایر توابع پیچیده‌تر است اما از حافظه‌ی کمتری استفاده می‌کند و در مسائلی که محدودیت در حافظه وجود دارد مناسب است. تابع لونبرگ[۳۰] به دلیل سریعتر بودن با استفاده از کاهش محاسبات، مورد استفاده قرار گرفته است. در اینجا شبکه عصبی وظیفه دارد تا در مرحله آموزش مدل با تعیین وزن‌ها، شرایطی مهیا سازد تا تشخیص روش در خصوص احساسات کاربران با مقدار واقعی آن کمترین اختلاف را داشته باشد. شبکه عصبی پس از

آموزش، جهت بررسی عملکرد با داده‌های تست ارزیابی می‌شود. در مدل پیاده‌سازی شده، شبکه عصبی شامل یک‌لایه ورودی ، پنج لایه نهان و یک‌لایه خروجی است.

مجموعه داده استفاده‌شده شامل ۱۴۶۴۰ توئیت در رابطه با خطوط هوایی ایالات متحده‌ی آمریکا است. این پایگاه داده از طریق سایت www.Kaggel.com قابل‌دسترسی است. هر توئیت به‌وسیله‌ی سه برچسب احساسی مثبت، منفی و خنثی برچسب‌گذاری شده است. شکل ۴ روند کلی روش پیشنهادی نهایی بهبود یافته را نمایش می‌دهد.



شکل۴: روند کلی روش ICM.

۷-معیارهای ارزیابی

معمولاً برای بررسی کارایی از معیار استاندارد دقت استفاده می‌شود. اگر تشخیص احساسات دو کاربر بر اساس متن توئیت آن‌ها به صورتی باشد که برای یک کاربر خاص، تشخیص صحیح و دیگری نادرست است، مقدار دقت طرح ۰/۵ (یا ۵۰ درصد) است. این امر نشان می‌دهد که مقدار بالاتر دقت تشخیص مطلوب‌تر را نشان می‌دهد. سایر معیارها نیز مانند صحت و فراخوانی جهت ارزیابی نتایج مورد استفاده قرار می‌گیرد. نحوه‌ی محاسبه‌ی هریک از معیارها در ادامه بیان‌شده است. رابطه ۹ نحوه‌ی محاسبه‌ی معیار دقت را نشان می‌دهد.

$$Precision = \frac{TP}{TP+FP} \quad (۹)$$

معیار دقت، نسبت تعداد «پیش‌بینی‌های صحیح انجام‌شده» برای نمونه‌های یک کلاس خاص، به تعداد «کل پیش‌بینی‌ها» برای نمونه‌های همان کلاس خاص را (این تعداد، مجموع تمامی پیش‌بینی‌های صحیح و پیش‌بینی‌های نادرست را شامل می‌شود) ارزیابی می‌کند. رابطه ۱۰ نحوه‌ی محاسبه‌ی معیار صحت را نشان می‌دهد.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (۱۰)$$

معیار صحت، بیان‌کننده تعداد «پیش‌بینی‌های صحیح انجام‌شده » توسط دسته‌بند، تقسیم بر تعداد «کل پیش‌بینی‌های انجام‌شده» توسط همان دسته‌بند است. رابطه۱۱ نحوه‌ی محاسبه‌ی معیار فراخوانی را نشان می‌دهد.

$$Recall = \frac{TP}{TP+FN} \quad (۱۱)$$

معیار فراخوانی، بیان‌کننده نسبت «تعداد داده‌های درست دسته‌بندی‌شده» در یک کلاس خاص، به تعداد کل داده‌هایی است که باید در همان کلاس خاص دسته‌بندی شوند، محاسبه می‌شود.

تعاریف مربوط به هریک از معیارها به شرح زیر است:

TP: بیانگر تعداد مواردی است واقعاً مثبت بوده‌اند و الگوریتم دسته آن‌ها را به‌درستی مثبت تشخیص داده است.

TN: بیانگر تعداد مواردی است که واقعاً منفی بوده‌اند و الگوریتم دسته آن‌ها را به‌درستی منفی تشخیص داده است.

FP: بیانگر تعداد مواردی است که واقعاً منفی بوده‌اند و الگوریتم دسته آن‌ها را به‌اشتباه مثبت تشخیص داده است.

FN: بیانگر تعداد مواردی است که واقعاً مثبت بوده‌اند و الگوریتم دسته آن‌ها را به‌اشتباه منفی تشخیص داده است.

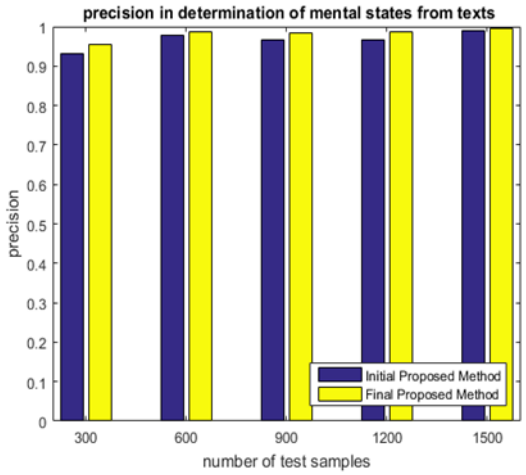
۸- ارزیابی نتایج

پس از آموزش شبکه، حال به ارزیابی روش IM، ICM و مقایسه روش ICM با سایر کارها پرداخته خواهد شد. برای ارزیابی طرح ۷۰ درصد پایگاه داده جهت آموزش، ۱۵ درصد جهت ارزیابی و ۱۵ درصد جهت تست مورد استفاده قرارگرفته است.

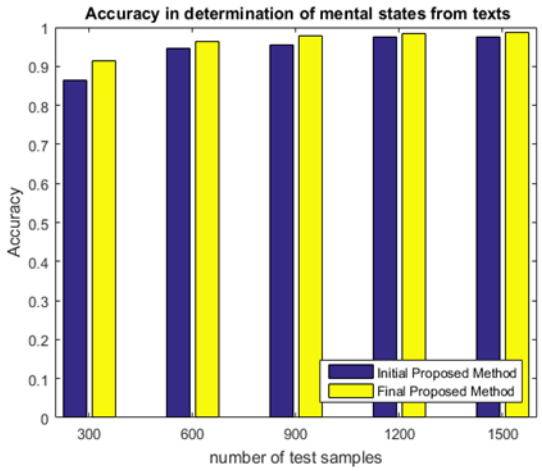
در شکل‌های ۵ تا ۷ دقت، صحت و فراخوانی روش ICM (استفاده از الگوریتم رقابت استعماری و شبکه‌ی عصبی پیچشی) با IM (استفاده از الگوریتم رقابت استعماری بدون در نظر گرفتن شبکه‌ی

عصبی پیچشی) جهت دسته‌بندی نمایش داده‌شده است. همان‌طور که در شکل می‌توان دید روش ICM دارای پارامترهای ارزیابی بهتری است.

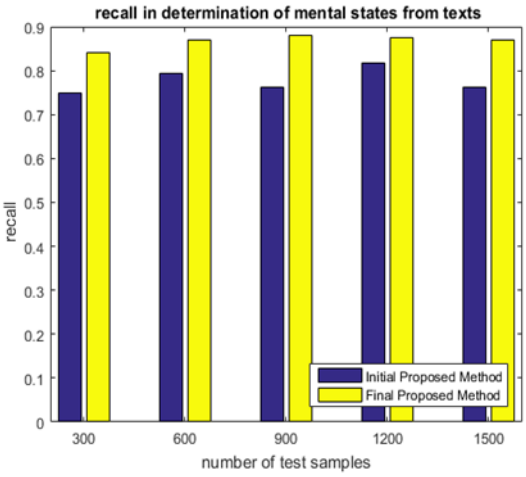
نتایج حاصل از مقایسه‌ها نشان‌دهنده‌ی تأثیر الگوریتم رقابت استعماری و شبکه‌ی عصبی پیچشی در افزایش دقت، صحت و فراخوانی است. همچنین در جدول ۷، دقت، صحت و فراخوانی روش ICM با سایر کارها مقایسه شده است که نتایج نشان‌دهنده‌ی مطلوب بودن روش ICM است. الگوریتم رقابت استعماری تاثیر به‌سزایی بر روی افزایش معیارهای ارزیابی دارد همچنین استفاده از شبکه عصبی پیچشی در کنار این الگوریتم به عنوان راهی موثر در جهت بهبود پارامترهای ارزیابی هم قابل توجه است.



شکل۵: مقایسه دقت روش IM با ICM.



شکل۶: مقایسه صحت IM با ICM.



شکل۷: مقایسه فراخوانی روش IM با ICM.

جدول۷: مقایسه روش ICM با سایر کارها.

ردیف	نویسندگان	دقت	صحت	فراخوانی
۱	Furqan Rustam, Imran Ashraf & et al [۲۴].	۰/۸۱۰	/۷۹۱	/۷۹۰
۲	Ankita Rane, Anand Kumar & et al [۲۵].	۰/۸۵۶	/۸۴۵	/۸۶۵
۳	Joshua Acosta, Norissa Lamaute & et al [۲۶].	/۷۵۰	/۷۲۰	/۷۲۰
۴	Yaser Maher Wazery, Hager Saleh Mohammed & et al [۲۷].	/۸۶۰	/۹۳۰	۰/۸۰
۵	**روش ICM**	**۰/۹۹۰**	**/۹۸۳۰**	**/۸۷۵**

۹- نتیجه‌گیری و کارهای آتی

هدف از این مقاله، ارائه روشی برای تشخیص احساس فرد بر اساس متن توئیت وی با استفاده از ترکیب الگوریتم فرا ابتکاری رقابت استعماری و شبکه‌ی عصبی پیچشی است به‌نحوی‌که بتواند با دقت بالا به تعیین حالت روحی فرد بپردازد. طرح پیشنهادی نهایی بهبود یافته (ICM) شامل سه مرحله است. ابتدا پیش‌پردازش متن جهت حذف کلمات اضافی انجام می‌شود. این کلمات به‌عنوان ورودی به روش بهینه‌سازی رقابت استعماری داده می‌شود تا زیرمجموعه‌ای از کلمات تأثیرگذار که در تعیین کلاس نمونه (احساس فرد) کمک می‌کند، تعیین شود. تابع برازش میزان مناسب بودن هر راه‌حل را بر اساس میزان همبستگی کلمات با برچسب نمونه و طول راه‌حل معین می‌سازد. در ادامه شبکه عصبی پیچشی برای یادگیری ویژگی به‌جای استخراج ویژگی مورد بهره‌برداری قرارگرفته است. گام بعد استفاده از یک دسته‌بند جهت طبقه‌بندی و پیش‌بینی دسته هر توئیت است. نتایج پیاده‌سازی حاکی از آن است که عملکرد طرح پیشنهادی نهایی بهبود یافته نسبت به سایر کارها نتایج بهتری داشته و مدل ارائه‌شده بر مبنای

آموزش ارائه‌شده به مدل، با دقت مطلوبی به تعیین وضعیت احساس فرد در زمان نگارش متن پرداخته است.

در تحقیقات بعدی هدف آن است که با سایر الگوریتم‌های فرا ابتکاری جدید مانند بهینه‌ساز شیر مورچه[۳۱] (ALO)، مدلی ارائه شود که با دقت بالا به تعیین احساسات فرد بر اساس توئیت‌های وی بپردازد. همچنین استفاده از ترکیب بهینه لایه‌های شبکه عصبی پیچشی به‌عنوان معماری مطلوب تشخیص احساسات بر اساس متن مورد بررسی و ارزیابی قرار خواهد گرفت. مورد دیگری که در تحقیقات بعد می تواند مورد ارزیابی قرار گیرد، بررسی انواع شبکه‌های عصبی مانند RNN [۳۲] در روش پیشنهادی است.

سپاسگزاری

در انتها بر خود لازم میدانیم که از سرکار خانم سارا خسروی، دانشجوی دکتری تخصصی دانشگاه رازی کرمانشاه، برای کمک ایشان در تهیه این مقاله، کمال تشکر و قدردانی را بعمل آوریم.

مراجع

[1] www.statista.com, https://www.statista.com/statistics/346167/facebook-global-dau, (2020).
[2] Binali, Haji, Chen Wu, and Vidyasagar Potdar. "Computational approaches for emotion detection in text." 4th IEEE International Conference on Digital Ecosystems and Technologies, IEEE, (2010).
[3] Neethu, M. S., and R. Rajasree. "Sentiment analysis in twitter using machine learning techniques." Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT). IEEE,(2013).
[4] Liu, Gang, and Jiabao Guo. "Bidirectional LSTM with attention mechanism and convolutional layer for text classification." Neurocomputing, (2019).
[5] Bandhakavi, Anil, et al. "Lexicon generation for emotion detection from text." IEEE intelligent systems 32.1, (2017).
[۶] سارا منتظریان، سید محمدحسین معطر. "تشخیص احساس متن فارسی با استفاده از روش مبتنی بر شباهت معنایی." کنگره بین‌المللی فن‌آوری، ارتباطات و دانش، (۱۳۹۴).
[7] Zobeidi, Shima, Marjan Naderan, and Seyyed Enayatallah Alavi. "Opinion mining in Persian language using a hybrid feature extraction approach based on convolutional neural network." Multimedia Tools and Applications 78.22 ,(2019).
[8] Zheng, Zhaohui, and Rohini Srihari. "Optimally combining positive and negative features for text categorization." ICML 2003 Workshop, (2003).
[9] Lucini, Filipe R., et al. "Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews." Journal of Air Transport Management 83, (2020).
[10] tan, Pang-Ning. Steinbach, Michael. Kumar, Introduction to data mining. Pearson Education, (2014).
[11] Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey." Ain Shams engineering journal 5.4, (2014).
[12] Ravi, Kumar, and Vadlamani Ravi. "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems 89, (2015).
[13] Mohammad, Saif M., and Peter D. Turney. "Crowdsourcing a word–emotion association lexicon." Computational Intelligence 29.3, (2013).
[14] Strapparava, Carlo, and Alessandro Valitutti. "Wordnet affect: an affective extension of wordnet." Lrec. Vol. 4. No, (2004).
[15] Ahmad, Munir, et al. "Tools and techniques for lexicon driven sentiment analysis: a review." Int. J. Multidiscip. Sci. Eng 8.1 , (2017).
[16] LeCun, Yann. "LeNet-5, convolutional neural networks". Retrieved 16 November, (2013).
[17] Zhang, Wei. "Shift-invariant pattern recognition neural network and its optical architecture". Proceedings of annual conference of the Japan Society of Applied Physics, (1988).
[18] Zhang, Wei. "Parallel distributed processing model with local space-invariant interconnections and its optical architecture". Applied Optics, (1990) .
[۱۹] فاطمه عباسی، بابک سهرابی، امیر مانیان، آمنه خدیور. "ارائه مدلی جهت دسته‌بندی احساسات خریداران کتاب با استفاده از رویکردی ترکیبی." فصلنامه مطالعات مدیریت کسب‌وکار هوشمند سال ششم، شماره ۲۱ ، (۱۳۹۶).
[20] E. Atashpaz-Gargari and C. Lucas, "Imperialist competitive algorithm: An algorithm for optimization inspired by imperialistic competition," 2007 IEEE Congress on Evolutionary Computation, Singapore, (2007).
[21] Qiancheng Fang, Hoang Nguyen, et al. " Prediction of Blast-Induced Ground Vibration in Open-pit Mines Using a New Technique Based on Imperialist Competitive Algorithm and M5Rules." Natural Resources Research, (2019).
[22] Vijayarani, S., and R. Janani. "Text mining: open source tokenization tools-an analysis." Advanced Computational Intelligence: An International Journal (ACII) 3.1, (2016).
[۲۳] نیکنام فرزاد و نیک نفس علی اکبر. "بهبود روش‌های متن‌کاوی در کاربرد پیش‌بینی بازار با استفاده از الگوریتم‌های انتخاب نمونه اولیه." مدیریت فناوری اطلاعات، دوره ۸، شماره۲, (۱۳۹۵).
[24] Rustam, Furqan, et al. "Tweets Classification on the Base of Sentiments for US Airline Companies." Entropy 21.11 (2019).
[25] Rane, Ankita, and Anand Kumar. "Sentiment classification system of twitter data for US airline service analysis." 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC). Vol. 1. IEEE, (2018).
[26] Acosta, Joshua, et al. "Sentiment analysis of twitter messages using word2vec." Proceedings of Student-Faculty Research Day, CSIS, Pace University 7, (2017).
[27] Wazery, Yaser Maher, Hager Saleh Mohammed, and Essam Halim Houssein. "Twitter Sentiment Analysis using Deep Neural Network." 2018 14th International Computer Engineering Conference (ICENCO).IEEE, (2018).

پاورقی‌ها:

[1] Imperialist Competitive Algorithm and Multilayer Perceptron network

[2] Imperialist Competitive Algorithm, convolution Network and Multilayer Perceptron Network

[3] Accuracy

[4] Naive Bayes

[5] Precision

[6] Long Short-Term Memory

[7] Attention-based Bidirectional long short term memory with convolution layer

[8] Domain-specific emotion lexicons

[9] General-purpose emotion lexicons

[10] Supervise

[11] Unsupervise

[12] Convolutional Neural Network

[13] Shift invariant

[14] Space invariant
[15] Imperialist Competitive Algorithm
[16] Bat Algorithm
[17] Imperialist
[18] Colony
[19] Integrated development and learning Environment
[20] Tokenization
[21] Stemming
[22] Stopwords
[23] Convolutional layer
[24] Pooling layer
[25] Fully-connected layer
[26] Rectified linear unit
[27] Multilayer perceptron
[28] Bayesian regularization backpropagation
[29] Scaled conjugate gradient backpropagation
[30] Levenberg_Marquardt backpropagation
[31] Ant lion
[32] Recurrent neural network