استخراج رابطه مبتنی بر تعبیه لغات با فرآیند جمع‌سپاری

محمد جعفرآباد* و روح‌الله دیانت

گروه مهندسی فناوری اطلاعات، دانشکده فنی و مهندسی، دانشگاه قم، قم، ایران

چکیده

برای انجام مطالعات داده‌کاوی، تاحدودی به‌دلیل پیچیده‌بودن فرآیند انتخاب ویژگی در کار مورد نظر، نیاز داریم تا بخشی از برچسب زنی را به کارگران در فعالیت جمع‌سپاری واگذار کنیم. فرآیند واگذاری کارهای داده‌کاوی به کاربران، اغلب به‌وسیلهٔ سامانه‌های نرم افزاری و بدون اطلاع دقیق از موقعیت سنی یا جغرافیای محل سکونت کاربران صورت می‌گیرد. عدم اطمینان از عملکرد کاربران مجازی در جمع‌سپاری، میزان صحت اطلاعات دریافتی را کاهش می‌دهد. در این مقاله پیشنهاد داده‌ایم تا با استفاده از روش‌های ایجاد انگیزش، تعدادی از مردم را در محلی جمع و از آنها در جهت وظایف جمع‌سپاری استفاده کنیم. افزایش دقت در اعلام نتایج به‌دلیل حضور فیزیکی، سرعت بالا در گرفتن نتایج با دقت بالا در زمان تعیین‌شده، تحصیلات مناسب شرکت‌کنندگان در فعالیت و بومی‌بودن طرح اجرایی از ویژگی‌های این پژوهش هستند. در این پژوهش یک کار یادگیری ماشین انجام شد تا بتوانیم در ضمن آن فعالیت‌های جمع‌سپاری را با الگوریتم‌های شبکه عصبی عمیق ترکیب کنیم. وظیفه رده‌بندی برای تعبیه لغات به‌صورت الگوریتمی و تلفیقی با کمک جمع‌سپاری انجام می‌شود. روش پیشنهادی با افزودن داده‌های جمع‌سپار به داده‌های قبلی و تغییرات در مدل تعبیه لغات ترکیبی گلاو و وردتووک توانست نتایج مناسبی را در استخراج ویژگی به‌دست بیاورد.

واژگان کلیدی: جمع‌سپاری، تعبیه لغات، گلاو، وردتووک، رده‌بندی

Relation extraction based on word embedding with Crowdsourcing Process

Mohammad Jafarabad* & Rouhollah Dianat

Department of Information Technology, Faculty of Engineering,
Qom university, Qom, Iran

Abstract

For data mining studies, due to the complexity of doing feature selection process in tasks by hand, we need to send some of labeling to the workers with crowdsourcing activities. The process of outsourcing data mining tasks to users is often handled by software systems without enough knowledge of the age or geography of the users' residence. We use convolutional neural network, for doing classification in six classes: USAGE, TOPIC, COMPARE, MODEL-FEATURE, RESULT and PART-WHOLE. This article extracts the data from the abstract of 450 scientific articles and it is a total of 835 relations. One hundred of these abstracts have been selected by the crowdsourcing. Classification results in this article have been done with a slight improvement in accuracy. In this study, we computed the classification results on a combination of vocabulary vectors with using of 450 abstract relation data (100 crowd source datasets with 350 standards). The results of the implementation of the classification algorithm give us performance improvement. This paper uses the population power to perform preparing data mining works. The proposed method by adding crowdsource data to the previous data was able to obtain better results rather than the top 5 methods.

Keywords: Glove, Word2vec, Crowdsourcing, word embedding, classification

* **Corresponding author**

*نویسندهٔ عهده‌دار مکاتبات

• تاریخ ارسال مقاله: ۱۳۹۸/۷/۵ • تاریخ پذیرش: ۱۳۹۹/۹/۱ • تاریخ انتشار: ۱۴۰۱/۰۳/۳۱ • نوع مطالعه: پژوهشی

[DOI: 10.52547/jsdp.19.1.101]

۱- مقدمه

هم‌گرایی فناوری‌ها این قابلیت را دارد تا چشم‌انداز جدیدی از همکاری را برای انجام امورات بشر باز کند. همانند جهش‌های قبلی در قدرت فناوری، هم‌گرایی محاسبات و ارتباطات اجتماعی، بهبود در زندگی را فراهم می‌کند. زیربناهای اقتصادی و علمی محاسبات فراگیر، چند دهه است که ساخته شده است. جهان فیزیکی، اجتماعی و مجازی، در حال تصادم، ادغام و توازن است. در جهان آینده، اعمال و کنش‌های ارتباطات و مجامع، متفاوت خواهند بود.

با پیشرفت فناوری، برخی از فعالیت‌های اجتماعی و اقتصادی با زیرساخت‌های فناوری اطلاعات ترکیب شده‌اند. دو اصطلاح جمعیت هوشمند و جمع‌سپاری نیز در این راستا ایجاد و تعریف می‌شوند. جمعیت هوشمند و فلش موب[1] که نخستین‌بار توسط هوارد رینگلد مطرح شد، به‌عنوان سازمان‌های اجتماعی خودساختار فناوری واسطه تعریف می‌شوند [1]. اصطلاح دیگر نیز جمع‌سپاری است که به عمل مشارکت گروهی مردم در به‌دست‌آوردن اطلاعات و انجام یک کار یا پروژه با استفاده از همکاری تعداد زیادی از مردم (با پرداخت یا بدون پرداخت) گفته می‌شود [2].

آیا می‌توان روشی را برای ایجاد یک جمع‌سپار پیشنهاد داد که به‌دلیل داشتن نوآوری و جذابیت، منتهی به جمعیت فعال مناسبی برای انجام فعالیت‌های جمع‌سپاری شود؟ مقاله [3] با مطالعات انجام‌شده به این نتیجه رسیده است که تنها دو مورد از نظریه‌های انگیزشی در مطالعاتشان در ایجاد جمع سپار بهره برده‌اند. همچنین مقاله [4] با بررسی مدل‌های نوآوری باز در حوزه جمع‌سپاری می‌نویسد: دانش و خلاقیت در نژاد بشر پراکنده شده‌اند و همیشه از توانایی‌های ما در استفاده از آن فراتر رفته‌اند. به طور سنتی، فرایندهای نوآورانه عمدتا (البته نه منحصرا) بر تلاش درون یک سازمان و یا همکاری بین سازمان‌های نسبتا "محدود" متکی هستند. به تازگی، مدل‌های نوآورانه "باز" توجه بیشتری را از سوی پژوهش‌گران و متخصصان نوآوری به‌عنوان روش‌های تلاش برای ادغام استعداد، دانش و ایده‌های توزیع شده در فرایندهای نوآورانه به سوی خود جلب کرده‌اند.

هر سیستم جدیدی که تعریف می‌شود، یک سری خدمات را ارائه می‌دهد که این خدمات یک راه حل برای انجام کارهایی هستند که درقبل به‌سادگی قابل حل نبوده‌اند. سامانه جمع‌سپاری راه حلی اجتماعی برای وظایف کوچکی که قبلا با برنامه‌نویسی قابل حل نبوده‌اند، ارائه می‌دهد. اما خود سیستم پیشنهادی نیز می‌تواند شامل ایراداتی باشد که در جمع‌سپاری بزرگترین مشکل تضمین کیفیت نتایج است. چگونه می‌توان به سمت مسیری حرکت کرد تا یک جمع‌سپار را با هزینه‌ای کمتر به مردم معرفی کرد؟ و یا وظایف جمع‌سپاری را در محیطی فیزیکی توسط جمعیت با قابلیت اطمینان بالا انجام داد؟

در ادامه مقاله به بررسی و تحلیل موضوع خواهیم پرداخت. در بخش دوم پیشینه تحقیق بیان می شود. در بخش سوم، مدل پیشنهادی شرح داده خواهد شد و در بخش چهارم مقایسه روش پیشنهادی با سایر روش‌ها طرح می‌شود. مقایسه رویکردی جمع‌سپاری، بحث و بررسی و نتیجه‌گیری بخش‌های پایانی این مقاله هستند.

۲- پیشینهٔ جمع‌سپاری

این مقاله ترکیب وظیفه‌های طبقه‌بندی مبتنی بر یادگیری عمیق با جمع‌سپاری است. لذا در این بخش ابتدا پیشینه‌ای از جمع‌سپاری بیان می‌شود و در ادامه پیشینه مقالات پایه یادگیری عمیق مورد استفاده در این پژوهش بیان می‌شود.

۲-۱- پیشینهٔ جمع‌سپاری

به‌طور سنتی، اصطلاح "جمعیت" تاحدودی به‌طور انحصاری در زمینه افرادی که در اطراف یک هدف، احساس یا تجربه مشترک هستند، استفاده می‌شود. با این حال امروزه اغلب شرکت‌ها به بحث در مورد چگونگی جمع‌آوری افراد برای اهداف سازمان می‌پردازند. جمع‌سپاری در [5] به‌عنوان استفاده از فناوری‌های اطلاعاتی برای اعطای مسئولیت‌های تجاری به جمعیت مورد استفاده قرار می‌گیرد. جمع‌سپاری یک راه مناسب برای حل مسئله است.

[1] فلش موب‌ها نوع خاصی از جمعیت‌های هوشمند می‌باشند که بر خلاف سایر گروه‌ها، اهدافی تفریحی و غیر معمول از تشکیل جمعیت، دارند. این درحالی است که سایر جمعیت‌های هوشمند افرادی هستند دارای اهداف محکم که جهت انجام کارهای گروهی مفید با استفاده از مزایای ارتباطات شبکه‌ای، گرد هم آمده‌اند.

[DOI: 10.52547/jsdp.19.1.101]

در سال‌های گذشته پردازش زبان طبیعی به سمت پژوهش‌هایی مثل سم ایول[1] و حاشیه‌نویسی رفته است. درهمین‌اواخر با ظهور پلتفرم‌هایی مثل ای ام تی[2] و جمعیت گل[3] و به‌طور هم‌زمان رشد درصد استفاده از اینترنت، جمع‌سپاری یکی از روش‌های سریع و ارزان برای پژوهش‌های پردازش متن و حاشیه‌نویسی شده است. جمع‌سپاری برای انجام ماشینی کارها به کار می‌رود، که در آن کارگران بابت کارهای انجام داده، پول دریافت می‌کنند. جمع‌سپاری کاربرد فراوانی در تحلیل احساسات دارد. یک نمونه از این فعالیت‌ها، برچسب‌زنی داده‌هاست، که کاربران برحسب حس خود، داده‌ها را برچسب‌زنی می‌کنند. برخی از پروژه‌ها در جمع‌سپاری با چندین زبان اجرا می‌شوند.

جمع‌سپاری توجه زیادی را در دنیای کسب و کار به ارمغان آورده است و بسیاری از شرکت‌ها ارزش کسب و کار بالقوه خود را با جمع‌سپاری به‌دست آورده‌اند و کارزارهایی را برای این اهداف راه‌اندازی کرده‌اند. در همین حال، جمع‌سپاری منحصر به اهداف تجاری نیست [3]. نیاز به نوآوری مداوم یکی از مهم‌ترین اولویت‌های کسب و کار در میان مدیران اجرایی و یک مسأله کلیدی در پژوهش‌های علمی است. با توجه به نیاز به جریان مستمر نوآوری در محصولات و ارائه خدمات جدید، شرکت‌ها به مخترعان حرفه‌ای برای تولید ایده‌ها متکی هستند. در همان مقاله، ۳۳ پژوهش جمع‌سپاری مطالعه شده است، تنها برخی از مطالعات بر روی یک پلتفرم خاص تمرکز کرده‌اند، مابقی به نوآوری، ارزیابی، حل مسأله، مدل، طراحی و اندازه‌گیری پرداخته‌اند.

تارنمای کیوا[4] [8] نمونه‌ای از جمعیت‌های هوشمند است که در پی آرمان‌های بشردوستانه پدید آمده است. در اینجا تمایل به برقراری ارتباط اساس تفکر پیوند مردم برای ایجاد روابط است. در این تارنما مردم می توانند به یکدیگر پول قرض بدهند و یا از یکدیگر قرض بگیرند. متوسط زمان برای پیداکردن وام در تارنمای کیوا ۹۶ ساعت است و بازپرداخت آن به‌صورت یکماهه، سه ماهه و یا در یک دوره زمانی محدود است.

به‌تازگی ایده‌هایی داده می‌شود که در بازاریابی از جمع‌سپاری استفاده شود [15]. ایده‌های بسیاری مطرح می‌شود، اما نبود بازاریابی مناسب منجر به شکست آنها می‌شود. برای این ایده‌ها هزینه هم پرداخت می‌کنند. تولیدکنندگان محصولات "فکر مشتری" را درک می‌کنند که این عمل بر مبنای ایده‌هایی است که نیازهای آنها را به‌طور مؤثرتر بررسی می‌کنند. در چالش بتا کاپ[5]، تنها ۷۰۰۰ یورو پاداش تعیین شد، این به غیر از هزینه‌هایی است که صرف تبلیغات و اطلاع‌رسانی این چالش شده است.

مقاله [3] آینده این پژوهش را مربوط به فعالیت‌هایی می داند که به بررسی عوامل پذیرش وظایف از طرف جمعیت دارند. استفاده مؤثر از جمعیت یکی از اهداف آینده این پژوهش‌ها است. این ایده قابلیت خواهد داشت تا در آینده با انجام یادگیری نیمه‌نظارت شده، به‌صورت مداوم به بهبود نتایج دسته‌بندی کمک کند. از جمله مسائل دیگری که می‌بایست در طرح‌های بعدی مورد لحاظ قرار بگیرد این است که با چه راه‌کارهایی می‌توان یک پایداری در روند مشارکت مردمی ایجاد کرد؟

پژوهش، طراحی و توسعه در دانشگاه بردلی (پوریا، نشان می‌دهد که یادگیری فعال یا یادگیری تجربی، یادگیری مبتنی بر بازی و گیمیفیکیشن، یادگیری مبتنی بر تجزیه و سایر یادگیری‌های مبتنی بر منابع انسانی و جمع‌سپاری می‌توانند کارایی بالایی داشته باشند [19].

مقاله [20] از چندین توانایی برچسب‌زن‌ها استفاده کرده است. این مقاله به بررسی روشی برای پیداکردن متخصصان در برچسب‌زنی می‌پردازد و با کمترین نرخ خطا بر روی پنجاه هزار مجموعه‌داده تصویری برچسب‌زنی را انجام داده است.

در مقاله [21] جمع‌سپاری در سندرم حاد تنفسی حاد کرونا مورد استفاده قرار گرفته است. نویسندگان با استفاده از جمع‌سپاری در شبکه‌های اجتماعی تصویری ارزشمند از شیوع کرونا در زمان واقعی ارائه می‌دهند؛ اما پوشش جغرافیایی ناهمگن است. در این روش استخراج اطلاعات برای پیش‌بینی و اطلاع‌رسانی استراتژی‌های شیوع و پس از آن تکرار این روند در چرخه‌های تکرارشونده برای نظارت و ارزیابی پیشرفت، وجود دارد. الگوریتم‌هایی نیز برای اعتبارسنجی نتایج ایجاد شده‌اند. مقاله [22] نیز بحث قیمت‌گذاری پویا را مورد بررسی قرار داده است. این مقاله همچنین بحث تجمیع و ارسال مجدد داده‌ها را برای جمع‌سپاری بررسی کرده است.

[1] SEMEVAL
[2] AMT
[3] Cf(Crowd flower)
[4] www.kiva.org - Loans that change lives

[5] Betacup: این مسابقه برای طرح ایده‌ای جهت کاهش تعداد فنجان غیر قابل بازیافت بود که هر ساله با ایجاد یک جایگزین مناسب تر برای فنجان قهوه قابل استفاده مجدد ایجاد می‌شد.

[DOI: 10.52547/jsdp.19.1.101]

مقاله [17] در سال ۲۰۱۶ شرکت‌کنندگان در جمع‌سپاری را اغلب افراد با اعتبار کم و نتایج آن را با کیفیت پایین گزارش می‌کند؛ لذا در بسیاری از موارد کشف حقیقت در نتایج جمع‌سپاری حاصل نمی‌شود. برخی از ویژگی‌ها که فعالیت جمع‌سپاری پیشنهادی ما را نسبت به کارهای انجام‌شده در پلتفرم‌هایی مثل آمازون و ای.تی.ام متمایز می‌کرد، عبارتند از: تضمین واقعی نتایج، گزارش در لحظه، استفاده از متخصصان واقعی، موقعیت جغرافیایی مشخص، کمک به انجام سریع، کمک به معرفی و ایجاد یک جمع‌سپار، ایجاد جریان تبلیغاتی برای فعالیت جمع‌سپاری، ایجاد ایده‌ای جذاب و بومی‌سازی جمع-سپاری.

مسأله‌ای که در مقاله [16] به آن اشاره شده است، این است که ایده جمع‌سپاری باید توجه رسانه‌ای را به خودش جلب کند. این مهم برای سمینار رقابت‌های استارت آپی اتفاق افتاد و اخبار آن در بیش‌تر رسانه‌ها، صدا و سیما و تارنماهای دانشگاه‌های معتبری مثل صنعتی شریف و امیرکبیر منتشر شد.

۲-۲- پیشینهٔ تعبیه لغات

زیروظیفهٔ یک، از وظیفهٔ هفتم سم ایول در سال ۲۰۱۸ به بررسی استخراج رابطه پرداخته است. در پژوهش حاضر همان زیروظیفه با استفاده از جمع‌سپاری انجام شده است؛ لذا دو مقاله [11] و [12] در این بخش به‌عنوان مقالات پایه بحث طبقه بندی مورد مطالعه قرار می‌گیرند. سایر مقالات این زیروظیفه در مقایسه‌های انجام‌شده در بخش نتایج آمده‌اند.

مقاله پایه [11] با موضوع استخراج رابطه در مقالات علمی، وظیفه استخراج رابطه معنایی و طبقه‌بندی در چکیده مقاله‌های علمی را تشریح می‌کند. این چالش بر روابط معنایی با دامنه خاص متمرکز است و شامل سه زیرشاخه مختلف است. زیرشاخه‌ها به‌گونه‌ای طراحی شده بودند که تأثیر مراحل مختلف پیش‌پردازش بر نتایج طبقه‌بندی رابطه را با یکدیگر مقایسه کنند. انتظار می‌رود که این وظیفه برای طیف گسترده‌ای از پژوهش‌گرانی که در استخراج دانش تخصصی فعالیت دارند، مرتبط باشد.

هدف از [11] این است که به‌طور خودکار روابط معنایی را در یک مقاله از نشریات علمی شناسایی کند. در این کار، جفت‌موجودیت‌ها به‌عنوان نمونه‌های ردهٔ ایستا لحاظ می‌شدند. این وظیفه شامل شناسایی و طبقه‌بندی نمونه‌های روابط معنایی بین مفاهیم در مجموعه‌ای از شش دسته گسسته است. این روابط خاص با حوزه علمی است و نمونه‌های آنها اغلب در چکیده یا مقدمه مقالات علمی یافت می‌شوند. این کار برای تهیه چارچوبی برای ارزیابی منظم مراحل لازم برای استخراج کامل اطلاعات از متن علمی انجام می‌شود.

مقاله پایه [12] نیز با موضوع طبقه‌بندی رابطه‌ای[1]، دو سامانه را برای طبقه‌بندی رابطه معنایی[2] توصیف کرده است. این مقاله وظیفه طبقه‌بندی رابطه معنایی را اجرا کرده است. یک مدل SVM[3] و یک مدل CNN[4] ویژگی‌های متراکم[5] وردتووک و ویژگی‌های پراکنده دستی را ترکیب می‌کنند. برای آموزش مدل‌ها، دو مجموعه‌داده ارائه شده برای زیرشاخه‌ها به‌منظور تعادل رده‌ها، ترکیب شدند. نتیجه این‌طور بود که عملکرد مدل SVM بهتر از CNN شد.

مقاله [6] به بررسی تجزیه و تحلیل احساسات اخبار با استفاده از جمع‌سپاری می‌پردازد. این مقاله رویکردی برای استخراج نمرات ۳۷۰۰۰ واژه داشته است. این رویکرد به‌وسیله تارنمای رپلر[6] پیاده‌سازی شده است. این مقاله از یک روش بدون نظارت یادگیری استفاده می‌کند. بسیاری از روش‌های تجزیه و تحلیل احساسات از منابع واژگانی استفاده می‌کنند. برخی مقالات لغت را خارج از متن بررسی کرده و می‌گویند هر لغت دارای یک قطبیعت مثب و منفی است. در مقاله [7] داده‌ها با چندین رده (از خیلی منفی تا خیلی مثبت) رده‌بندی شده‌اند. که رتبه‌بندی این داده‌ها به‌وسیلهٔ جمع‌سپاری انجام شده است.

۳- روش پژوهش

استخراج رابطه، کاری برای استخراج اطلاعات است که هدف آن شناسایی و طبقه‌بندی روابط معنایی بین موجودیت‌ها در متن است. این کار به‌طوراساسی ساختار را از اطلاعات متنی بدون ساختار استخراج می‌کند و به ما این امکان را می‌دهد تا اطلاعات ارزشمندی را در مورد نحوه تعامل موجودیت‌ها به‌دست آوریم. در نتیجه ظرفیت انسانی را برای تجزیه و تحلیل (اغلب زیاد) داده‌های متنی بهبود می‌بخشد. در استخراج، جفت‌موجودیت‌ها[7] از یک سند بررسی و نوع رابطه پیش‌بینی می‌شود.

[1] relation classification
[2] semantic relation classification
[3] ماشین بردار پشتیبان
[4] convolutional neural network
[5] word2vec
[6] rappler.com
[7] pairs of entities

[DOI: 10.52547/jsdp.19.1.101]

مطالب علمی با ساختارهای پیچیده و تکیه بر اصطلاحات تخصصی پیش‌بینی نوع رابطه صحیح[1] را دشوارتر می‌کنند. در پژوهش پیش روی مشکل تشخیص رابطه به‌عنوان یک کار طبقه‌بندی با ناظر کار شده است. در اینجا ناظر جمع‌سپاری است.

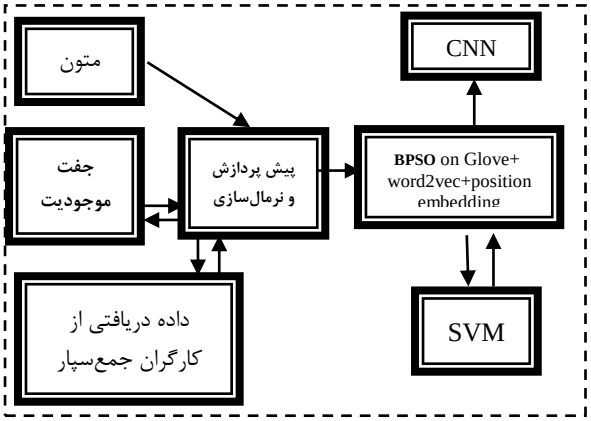
۱-۳- مدل طرح پیشنهادی

دو مدل تحلیل شده‌اند. یک ماشین بردار پشتیبانی (SVM)، که از مجموعه‌ای غنی از ویژگی‌ها در ترکیب ویژگی‌های متراکم وردتووک و ویژگی‌های پراکنده دستی بهره برده است و در روشی دیگر از یک مدل عصبی پیچشی (CNN) استفاده می‌شود. بهترین نتیجه با مدل SVM به‌دست آمد.

روش پیشنهادی ما برای استخراج رابطه معنایی شامل سه مرحله اساسی است: تهیه داده‌ها، تصمیم‌گیری در مورد ورودی‌ها و آموزش مدل CNN. شکل (۱) نمای کلی مدل پیشنهادی را نشان می‌دهد. ما ابتدا بر روی اسناد ورودی کار می‌کنیم تا برخی از اقدامات پیش پردازش از جمله نشانه‌گذاری را انجام دهیم. ما به دنبال بردارهای تعبیه لغات هستیم؛ لذا از مدل‌های پیش‌آموزش داده‌شده وردتوورک و گلاو[2][28] بر روی مجموعه‌متن‌های بزرگ استفاده کرده‌ایم. همچنین تعبیه موقعیت برای هر نشان در یک جمله معین می‌شود.

در پایان مرحله پیش‌پردازش و در مرحله آموزش، هر سه بردار جاسازی‌شده با هم جمع می‌شوند تا یک ماتریس ورودی اصلی با اندازه n * (e1 + e2 + d) ایجاد شود که در آن n بیشینه طول جملات نشانه‌گذاری شده است وe1 و e2 به‌ترتیب طول بردارهای تعبیه‌شده به‌ازای هر نشانه از وردتووک و گلاو هستند. طول تعبیه موقعیت برای هر کلمه به‌وسیله d نشان داده شده است.

(شکل-۱): مدل پیشنهادی برای طبقه‌بندی موجودیت‌ها
(Figure-1): Proposed model for classifying entities



ممکن است در جملات، متن چندکلمه‌ای وجود داشته باشد که در یک تعبیه یا در هر دو مدل از قبل آموزش دیده نشده باشند. یک بردار صفر برای این کلمات استفاده می‌شود. برای هر مدل تعبیه‌ای که آنها را پوشش نمی‌دهد، این بردار لحاظ می‌شود. آخرین عملیاتی که در مورد بردارهای تعبیه‌شده انجام می‌دهیم، نرمال‌کردن آنها به مقادیر مشابه است.

۲-۳- نوآوری مدل پیشنهادی

در مرحله بعد از آماده‌سازی داده‌ها، زیرمجموعه‌ای کارآمد از ویژگی‌های ورودی را با استفاده از BPSO[3] انتخاب می‌کنیم. برای اعمال BPSO، باید یک تابع برازش[4] تعریف و به‌طور مشابه مقادیر مناسبی را برای هایپر پارامتر الگوریتم تعریف کنیم. نمره F به‌دست می‌آید که با استفاده از الگوی آموزش‌دیده با هر زیرمجموعه انتخابی از ورودی‌ها در مجموعه‌داده‌های توسعه به‌عنوان مقدار برازش برای BPSO در نظر گرفته می‌شود.

بهینه‌سازی ازدحام ذرات بدودویی یا همان BPSO به یک تابع برازش[5] نیاز دارد تا میزان برتری راه حل انتخابی را به‌دست بیاورد؛ لذا می‌بایست یک طبقه‌بند با ویژگی‌های انتخاب‌شده برای هر بردار ایجاد شود. این طبقه‌بند بر روی داده توسعه‌ای به‌وسیلهٔ SVM اعمال می شود. این در حالی است که در سایر مدل‌ها با CNN انجام می‌شد، که اگرچه دقت بیشتری داشت، ولی زمان بیشتری برای به نتیجه‌رسیدن نیاز داشت. طول دوره آموزش در CNN زیاد بود؛ لذا طبقه‌بند به SVM در BPSO تغییر داده شد؛ لذا با توجه به تعداد بسیار زیاد پارامترها لازم است که مدل برای CNN آموزش داده شود و تصمیم گرفتیم از یک فرآیند انتخاب ورودی نیز به جای استفاده از CNN استفاده کنیم. دلایل اصلی انتخاب SVM عبارتند از: تعداد بسیار کم هایپر پارامترهای قابل تنظیم و توانایی آن برای کار با تعداد زیادی از ویژگی‌های ورودی نیز بالا است. گفتنی است که ما به هایپرپارامترهای SVM[6] اهمیت نمی‌دهیم و مقادیر پیش‌فرض را بدون تغییر نگه می‌داریم.

[3] binary particle swarm optimization

[4] Fitness Function برای اینکه خوب یا بدبودن جواب را تعیین کنیم، از مفهومی به نام تابع برازش استفاده می‌کنیم. اگر تابع نتواند میزان خوب‌بودن جواب را به‌درستی نشان دهد، الگوریتم معیار ارزیابی دقت عملکرد نخواهد داشت.

[5] fitness function

[6] support vector machine

[DOI: 10.52547/jsdp.19.1.101]

بعد از همه این مراحل، الگوی آموزش‌دیده با CNN برای داده‌های آزمایش به‌کار می‌رود تا روابط معنایی بین موجودیت ذکر شده در جملات داده‌شده را پیش‌بینی کند. در مرحله تست، پس از آماده‌سازی بردارهای تعبیه‌شده کلی، ما فقط از مقادیر ابعادی که به‌وسیلهٔ BPSO در مرحله آموزش تعیین می‌شوند، استفاده می‌کنیم.

تغییر الگوریتم از PSO به BPSO، استفاده از SVM ، تعبیه موقعیت و جمع‌سپار با هدف بهبود نتایج با سرعت بیشتر نوآوری این پژوهش است. همچنین در این پژوهش از بسیاری از روش‌های استخراج ویژگی اضافی مورداستفاده در سایر مدل‌ها استفاده نشده است؛ درنهایت با این تغییرات سامانه پیشنهادی در بین پنج سامانه برتر قرار گرفت که نباید در رسیدن به این نتیجه نقش جمع‌سپاری و مراحل آن را نادیده گرفت.

۳-۳- مجموعه‌داده

برای مجموعه‌داده از سم ایول ۲۰۱۸ استفاده می‌شود. با توجه به اینکه در آنجا یک مجموعه ثابت لحاظ شده است و همه مقالات بر اساس آن نتایج را ارزیابی و اعلام کرده‌اند، ما نیز در مرحله نخست تنها به همین مجموعه داده بسنده کرده‌ایم. اما در ادامه داده‌های جمع‌سپاری را هم برای دقت عملکرد بیشتر اضافه کرده‌ایم؛ لذا ما در مرحله نخست فقط ترکیب الگوریتمی را تغییر داده‌ایم؛ اما در مرحله دوم با اضافه‌شدن نتایج صد چکیده مطلب دریافتی از جمعیت جمع‌سپار توانستیم تا دقت را افزایش دهیم. داده جدید از متن صد چکیده مرتبط با موضوعات علمی عبارت "Extraction Relation" در سایت IEEE انتخاب شد.

مجموعه‌داده مورد استفاده مطابق شکل (۲)، یک سری چکیده در حوزه متن‌کاوی بودند که به‌صورت زوج‌واژه انتخاب و برچسب‌گذاری بر روی آنها انجام شده بود. به‌عنوان مثال در شکل (۲) که بخشی از کل مجموعه‌داده است، چکیده مطلب با شماره H01-1001 برای سه مورد زوج واژه یافت می‌شود. که در سه ردهٔ مختلف قرار دارند. این مجموعه‌داده شامل چکیده مطلب و انتخاب موجودیت در آنها است.

۱-۳-۳- جمعیت هوشمند طرح پیشنهادی

بیش‌تر جوامع جمع تشکیل‌شده سابقه طولانی ندارند؛ بنابراین هیچ سابقه‌ای از موفقیت و شکست در این زمینه وجود ندارد. مقاله [16] از نظرات شرکت‌کنندگان در کارگاه‌های پژوهشی در دانشگاه کارولینای شمالی، دانشکده وارتون در دانشگاه پنسیلوانیا، دانشگاه ایالتی کارولینای شمالی، دانشگاه کارولینای جنوبی، دانشگاه ییل، دانشگاه سانتا کلارا، دانشگاه تمپل، دانشگاه هاروارد و دانشگاه جنوبی کالیفرنیا در طول تکمیل پژوهش استفاده کرد. در پژوهش حاضر انجام مطالعات در دانشگاه شهید شمسی‌پور و دانشگاه امیرکبیر انجام شده است.

اگرچه وجود عوامل کلیدی عنصر لازم برای هر جمعیت هوشمند است، اما علاوه‌بر عوامل کلیدی گفته‌شده، یک‌سری پارمترهای اختصاصی نیز در تشکیل جمعیت هوشمند پیشنهادی مؤثر بودند. در جدول (۱) به کلیه پارامترهای اختصاصی اشاره کرده‌ایم.

(شکل-۲): مجموعه‌داده برچسب زده‌شده

(Figure-2): labeled datasets

H01-1001	H01-1001.5	H01-1001.7	USAGE-REVERSE
H01-1001	H01-1001.9	H01-1001.10	USAGE
H01-1001	H01-1001.14	H01-1001.15	PART_WHOLE-REVERSE
H01-1017	H01-1017.4	H01-1017.5	MODEL-FEATURE
H01-1041	H01-1041.3	H01-1041.4	PART_WHOLE-REVERSE
H01-1041	H01-1041.8	H01-1041.9	USAGE
H01-1041	H01-1041.10	H01-1041.11	MODEL-FEATURE-REVERSE
H01-1041	H01-1041.14	H01-1041.15	USAGE-REVERSE
H01-1042	H01-1042.1	H01-1042.3	USAGE
H01-1042	H01-1042.10	H01-1042.11	MODEL-FEATURE
H01-1049	H01-1049.3	H01-1049.4	PART_WHOLE-REVERSE
H01-1058	H01-1058.2	H01-1058.4	RESULT
H01-1058	H01-1058.9	H01-1058.10	RESULT
H01-1058	H01-1058.13	H01-1058.14	RESULT-REVERSE
H01-1058	H01-1058.16	H01-1058.18	USAGE-REVERSE
H01-1058	H01-1058.19	H01-1058.20	RESULT
H01-1058	H01-1058.24	H01-1058.25	MODEL-FEATURE-REVERSE
H01-1058	H01-1058.27	H01-1058.28	MODEL-FEATURE-REVERSE

(جدول-۱): عوامل خاص منظوره برای ایجاد جمعیت هوشمند

(Table-1): Special-purpose factors to create crowdsourcing

عامل اختصاصی جمعیت هوشمند پیشنهادی	توضیحات
محدوده جغرافیایی کمپین	اگر جمعیت هوشمند در محلی با دسترسی عمومی آسان ایجاد شود، امکان حضور بیشتری در محل وجود دارد، در این پژوهش دانشگاه شهید شمسی‌پور در فاصله ۷۰۰ متری از میدان امام حسین (ع) که مرکز تجمعات استان تهران است، جهت ایجاد جمعیت انتخاب شد.
بانک اطلاعاتی مرتبط	استخراج خودکار از وب‌سایت‌های مرتبط با کسب و کار
برقراری اعتماد	هسته اصلی جمعیت‌های هوشمند اعتماد است. برگزاری یک رویداد در دانشگاه این اعتماد را تا حد بالایی برقرار می‌سازد.
اهمیت محتوا در اطلاع رسانی پیامکی	هر پیام متنی به تنهایی در هر گوشی تلفن همراه می تواند به‌عنوان عموم مردم شناخته شود. نظر به اینکه که هر فرد قدرت و قابلیت نوشتن، فرستادن و دوباره فرستادن پیام‌ها را دارد، اثر این امر بسیار پر اهمیت و قابل توجه است.
ایجاد حس رقابت	رویداد جمعیت هوشمند مد نظر با عنوان رقابت های استارت آپی و با دعوت از رقبای حاضر در حوزه‌های شغلی مختلف، توانست تا جمعیت زیادی را برای مشاهده مناظره شغلی گرد هم جمع کند. مردم تمایل زیادی به تماشای رقابت‌های کسب و کاری دارند.

[DOI: 10.52547/jsdp.19.1.101]

استفاده از پتانسیل اتحادیه‌های شغلی	رویداد جمعیت هوشمند در هر جلسه یک حوزه شغلی را مورد مطالعه قرار داد. اتحادیه‌ها و اصناف کمک به‌سزایی در ارسال پیامک برای مخاطبین حوزه‌های شغلی خود داشتند و بدین شکل جمعیت هوشمند هر جلسه تا حدی نیز تخصصی و با حضور اعضاء آن صنف تشکیل شد. ارسال پیامک تبلیغاتی به گروه‌های هدف نیز تأثیر به‌سزایی در افزایش جمعیت همگون داشت. برای مثال در رقابت‌های استارت آپی املاک به ۱۳ هزار نفر از مشاوران املاک استان تهران پیامک ارسال شد.
ایجاد حس نوآوری	مردم علاقه‌مند به کسب اطلاعات نوین در حوزه کسب و کار هستند. از طرفی میل ذاتی انسان به فراگیری، کشف و تحصیل رضایت می‌تواند با تحول و رشد جمعیت‌های هوشمند در دهه اخیر مرتبط باشد. میل به کسب دانش و اطلاعات نوین نقش مؤثری در تشکیل این جمعیت هوشمند داشت. با توجه به اینکه ذات استارت‌آپ‌ها، شامل مفاهیم خلاقیت و ایده‌پردازی است، استفاده از آنها کمک به‌سزایی در تشکیل جمعیت دارد.
انجام تبلیغات انگیزشی صوتی/تصویری	ایجاد پادکست‌های تبلیغاتی از استارت آپ ها و ساخت کلیپ تبلیغاتی از استارت‌آپ‌ها
اجرا در محیط دانشگاهی	بنابر مقاله [16] از مطالعه میان ۵۵ نشریه، ۳۵ مقاله (۶۴٪) تنها توسط دانشگاهیان، ۱۰ مقاله (۱۸٪) تنها از صنعت و ۱۰ مقاله (۱۸٪) توسط همکاری دانشگاهیان و صنعت نوشته شده است. این نشان می‌دهد که دانشگاهی‌ها نقش مهمی در تحقق اهداف پژوهش دارند.

در طرح پیش رو جمعیتی حدود ۵۵۰ نفر از صاحبان کسب و کارها، مدیران کانال‌ها و گروه‌ها در شبکه‌های اجتماعی و علاقه‌مندان به توسعه کسب و کار در همایش رایگان رقابت‌های استارت آپی حاضر شدند، این جمعیت در سمینار به‌طور اشتراکی موجودیت‌های صد چکیده مطلب را برچسب‌زنی کردند.

در طرح پیشنهادی جمعیتی شامل فارغ‌التحصیل و دانشجو در بیش از پنجاه دانشگاه در فرآیند جمع‌سپاری شرکت کرده بودند، که حدود سی درصد از دانشگاه‌های دولتی ایران و نزدیک به یک درصد جمعیت نیز از دانشگاه‌های خارج از ایران بودند. بیش از هشتاد درصد شرکت‌کنندگان تحصیلات دانشگاهی داشتند و در حدود ۳۸ درصد از شرکت‌کنندگان در همایش تحصیلات کارشناسی ارشد به بالا داشتند.

۴- سناریوی پیشنهادی تعبیه برای استخراج رابطه با داده جمع‌سپار

سناریوهای مختلفی برای فعال‌کردن جمع‌سپاری می‌توان نوشت؛ به‌عنوان مثال در [10] بر اساس سناریو پاسخ‌های کم و پاداش کم پژوهش انجام گرفته است. نکته‌ای که می‌بایست در فعالیت‌های جمع‌سپاری به آن توجه شود این است که مردم در انجام فعالیت‌های جمع‌سپاری احساس خستگی نکنند. بهترین سناریوها، آنهایی خواهند بود که از تأثیرگذاری مردم بر مردم در جهت انجام وظایف کمک بگیرند و در راستای فعالیت‌های مثبت و ارزنده اجتماعی تلقی شوند. سناریونویسی از مهم‌ترین بخش‌های مدیریت جمعیت‌های هوشمند است.

در این پژوهش با استفاده از اطلاع‌رسانی پیامکی، تبلیغات در شبکه‌های اجتماعی و قدرت خبری رسانه‌ای، یک رویداد کسب و کار به نام "رقابت‌های استارت آپی" به‌وسیلهٔ "مجله صدای زنده" در دانشگاه شمسی‌پور تهران تشکیل شد. در هر جلسه از این رویداد، یک حوزه کسب و کاری مورد مطالعه قرار می‌گیرد. با توجه به مطالعات انجام‌شده، در کشور ایران استارت آپ‌ها در حوزه‌های مختلف شغلی فعالیت می‌کنند که از این بین در ۴۲ مورد از آن حوزه‌ها، رقبای استارت آپی وجود دارند؛ به‌عنوان مثال ۵۶۰ تاکسی برخط، ۵ آزمایشگاه پزشکی برخط، ۱۴ معماری برخط و ۱۲ خشک‌شویی آنلاین در ایران مشغول به فعالیت هستند. این رویداد پس از یازده جلسه به جهت برندسازی به دانشگاه امیرکبیر منتقل شد.

در سناریوی پیشنهادی تصمیم گرفتیم تا برای ایجاد جمعیتی با سطح سواد بالا جهت مطالعات متن‌کاوی آینده، همایش آموزشی رایگان رقابت‌های استارت آپی را با حضور مردم برگزار کنیم. بنا شد تا در زمان برگزاری این همایش از مردم در جهت فعالیت جمع‌سپاری کمک بگیریم. هزینه‌ای به‌عنوان جایزه معادل صد هزار تومان برای دو نفر به‌طور قرعه در نظر گرفته شد. هزینه‌های اجرایی برگزاری جلسه و تبلیغات برای جمع‌آوری این جمعیت نیز از جمله هزینه‌های این فرآیند جمع‌سپاری بود.

۴-۱- وظیفهٔ رده‌بندی در تعبیه لغات ترکیبی به‌وسیلهٔ فرآیند جمع‌سپار

مقاله [11] با استفاده از شبکه عصبی پیچشی[1] به‌عنوان یک وظیفه، رده‌بندی را برای شش ردهٔ کاربرد[2]، نتیجه‌گیری[3]، جزء-کل[1]، موضوع[2]، مدل-ویژگی[3] و

[1] Convolutional Neural Network
[2] USAGE
[3] RESULT

[DOI: 10.52547/jsdp.19.1.101]

مقایسه[4] تشخیص می‌دهد. این مقاله داده مورد نیاز خود را از چکیده ۳۵۰ مقاله علمی استخراج کرده است. از این تعداد چکیده درجمع ۸۳۵ رابطه استخراج شد.

در جدول (۲) این رده‌هاها بر اساس تعداد در هر رده و همچنین نوع رابطه مشخص شده‌اند. به‌عنوان مثال جمله "حافظه فیزیکی بخشی از رایانه است" یک رابطه جزء به کل عادی است و جمله "رایانه دارای مادربورد است" یک رابطه جزء به کل معکوس است. ترتیب آمدن موجودیت‌ها در جمله، جزء به کل عادی یا جزء به کل معکوس بودن را مشخص می‌کنند.

(جدول-۲): ردهٔ توزیع روی داده‌های یادگیری [11]
(Table-2): Distribution class on learning data[11]

نوع رابطه	عادی	معکوس	کل
کاربرد	296	187	483
موضوع	8	10	18
مقایسه	95	-	95
مدل-ویژگی	226	100	326
نتیجه‌گیری	52	20	72
جزء-کل	158	76	234
مجموع	835	393	1228

(جدول-۳): نتایج تفصیلی مبتنی بر اندازه‌گیری در مرحله انتخاب ویژگی ورودی. هر ردیف طبقه‌بند آموزش داده شده با SVM را با استفاده از ویژگی‌های ورودی یادشده نشان می‌دهد.
(Table-3): Detailed measurement-based results in the input feature selection step. Each row shows SVM-trained classifiers using the listed input features.

	دقت[5]	فراخوانی[6]	امتیاز F[7]	طول ورودی
W2vec	68.4	69.6	68.9	300
گلاو	67.2	68.4	67.7	300
تعبیه موقعیت	21.3	23.6	22.3	50
W2vec + گلاو	70.2	71.6	70.8	600
W2vec + گلاو+ تعبیه موقعیت	69.3	71.6	70.4	650
ورودی های نخبه	73.4	76.2	74.7	368

گلاو[8] و وردتووک[9] دو مورد از محبوب‌ترین الگوریتم‌های تعبیه لغات هستند. در کاربردهای مختلف یادگیری ماشین از این دو الگوریتم به‌طور مجزا استفاده شده است. ما با استفاده از مقاله [12] و با ترکیب گلاو و وردتووک بردار تعبیه‌ای جدید برای لغات ایجاد کرده‌ایم. این تغییر در جهت تکمیل وظیفه سم ایول سال ۲۰۱۸ در جهت ارزیابی با هدف کشف ماهیت معانی نوشته شده است، ابتدا دو بردار گلاو و وردتووک، هر یک به طول سیصد برای کلیه کلمات استخراج شده و سپس با اضافه‌شدن پنجاه بیت برای تعبیه موقعیت لغات، کلیه لغات به‌صورت برداری به‌طول ۶۵۰ تبدیل می‌شوند؛ در نهایت جملات براساس ترکیب لغات ساخته می‌شوند. نتایج اولیه تغییرات مدل(الگوریتم) در جدول (۳) نشان داده شده است.

برای تغییر در مقاله [12] از کدهای CNN و SVM همراه با BPSO استفاده شده است؛ به‌نحوی که SVM برای برازش تابع BPSO خواهد بود. جدول (۳) نتیجه استفاده از ورودی‌های نخبه را نشان می‌دهد.

حال می‌بایست از ورودی‌های نخبه در استخراج رابطه استفاده کنیم. از نتایج به‌دست‌آمده حاصل از بردارهای لغات نخبه، زوج موجودیت‌های متن مطابق جدول (۲) با CNN پیش‌بینی می‌شوند. بهبود در جدول (۴) آمده است.

(جدول-۴): مقایسه رده‌بندی‌های مبتنی بر تلفیق و داده استاندارد با حالت اضافه‌شدن داده جمع‌سپاری
(Table-4): Comparison of standard based classifications and data aggregation with additional data supplied by crowdsourcing

رده‌بندی در	امتیاز F[12]	فراخوانی[11]	دقت[10]
ETH-DS3Lab [13]	81.7	84.4	79.2
UWNLP [14]	78.9	-	-
SIRUS_LTG_UiO [23]	76.7	-	-
ClaiRE [24]	74.9	-	-
Talla [25]	74.2	-	-
MIT-MEDG [26]	72.7	-	-
TakeLab [27]	69.7	75.57	64.64
تعبیه مبتنی بر گلاو و وردتووک با ۳۵۰ چکیده اولیه و ۱۰۰چکیده جمع سپار	71.20	72.80	70.20

۵- نتیجه‌گیری

جمعیت هوشمند می‌تواند متخصص یا غیرمتخصص باشد. در جمعیت هوشمند غیر متخصص افراد از رده‌های سنی مختلفی حضور دارند. فعالیت حاضر با جمعیت

[1] PART-WHOLE
[2] TOPIC
[3] MODEL-FEATURE
[4] COMPARE
[5] Precision
[6] Recall
[7] F-score
[8] https://nlp.stanford.edu/projects/glove/
[9] http://code.google.com/archive/p/word2vec
[10] Precision
[11] Recall
[12] F-score

[DOI: 10.52547/jsdp.19.1.101]

به‌طورتقریبی متخصص انجام شد. همچنین می‌توانند از اعضای داخل سازمان یا غیر وابسته به سازمان باشند که بیش‌تر جمعیت حاضر از افراد غیر وابسته به سازمان بودند. همچنین نوع فعالیت می‌توانست فعالیت وابسته به یکدیگر در جمعیت یا فعالیت مستقل باشد در این پژوهش راه حلی نوین برای بومی‌سازی جمع‌آوری داده‌ها و کمک به مسائل داده‌کاوی با کمترین هزینه و کمک جمعیت متخصص ارائه شد. استراتژی‌های ابتکاری آموزش و یادگیری، که از فن‌آوری کمک می‌گیرند به راه‌کارهای نوینی در تجزیه و تحلیل داده تبدیل شده‌اند.

در این مقاله با استفاده از BPSO و SVM بردار تعبیه بهینه اشتراکی گلاو و ورد۲وک انتخاب شد. در پژوهش‌های قبلی این بردار مبتنی بر PSO و CNN بود؛ سپس با استفاده از CNN استخراج ویژگی موجودیت‌ها در چکیده متون علمی انجام شد، در این پژوهش با کمک ۴۵۰ داده (صد مجموعه‌داده جمع‌سپار به‌همراه ۳۵۰ مجموعه استاندارد قبلی) دقت به‌نحوی افزایش پیدا کرد که روش پیشنهادی ضمن افزایش سرعت، در بین پنج روش برتر قرار گرفت.

۸- مراجع 8- References

[1] H. Rheingold, Smart mobs: The next social revolution. Basic books, 2007.

[2] D. Zhou, Q. Liu, J. C. Platt, C. Meek, & N. B. Shah, Regularized minimax conditional entropy for crowdsourcing. arXiv preprint arXiv:1503.07240, 2015.

[3] Y. Zhao, & Q. Zhu, "Evaluation on crowdsourcing research: Current status and future direction", *Information Systems Frontiers*, vol. 16(3), pp. 417-434, 2014.

[4] S. Marjanovic, C. Fry, & J. Chataway, "Crowdsourcing based business models: In search of evidence for innovation 2.0", *Science and public policy*, vol. 39(3), pp. 318-332, 2012.

[5] J. Prpić, P. P.Shukla, J. H. Kietzmann, & I. P. McCarthy, "How to work a crowd: Developing crowd capital through crowdsourcing", *Business Horizons*, vol. 58(1), pp. 77-85, 2015.

[6] J. Staiano and M. Guerini, "DepecheMood: a Lexicon for emotion analysis from crowd-annotated news," *arXiv preprint arXiv*1405, pp. 1605, 2014.

[7] P. Gonçalves, M. Araújo, F. Benevenuto, and M. Cha, "Comparing and combining sentiment analysis methods," in Proceedings of the first ACM conference on Online social networks, 2013, pp. 27-38.

[8] P. Belleflamme, T. Lambert, & A. Schwienbacher, "Crowdfunding: Tapping the right crowd", *Journal of business venturing*, vol. 29(5), pp. 585-609, 2014.

[9] J. Daniels, & J. R. Feagin, "The (coming) social media revolution in the academy," *Fast Capitalism*, vol.8(2), 2019.

[10] T. A. Gautre, & T. H. Khan, "An analysis of question answering system for education empowered by crowdsourcing" In 2018 2nd International Conference on Inventive Systems and Control (ICISC), IEEE, 2018.

[11] K. Gábor, D. Buscaldi, A. K. Schumann, B. QasemiZadeh, H. Zargayouna, & T. Charnois, "Semeval-2018 Task 7: Semantic relation extraction and classification in scientific papers," In Proceedings of The 12th International Workshop on Semantic Evaluation, pp. 679-688, 2018.

[12] M. Gluhak, M. P. di Buono, A. Akkasi, & J. Šnajder, "TakeLab at SemEval-2018 Task 7: Combining Sparse and Dense Features for Relation Classification in Scientific Texts", In Proceedings of The 12th International Workshop on Semantic Evaluation, pp. 842-847, 2018.

[13] J. Rotsztejn, N. Hollenstein, & C. Zhang, Eth-ds3lab at semeval-2018 task 7: Effectively combining recurrent and convolutional neural networks for relation classification and extraction. arXiv preprint arXiv:1804.02042, 2018.

[14] Y. Luan, M. Ostendorf, & H. Hajishirzi, "The uwnlp system at semeval-2018 task 7: Neural relation extraction model with selectively incorporated concept embeddings", In Proceedings of The 12th International Workshop on Semantic Evaluation, pp. 788-792, 2018, June.

[15] S. A. Lazarus, Cyber Mobs: A Model for Improving Protections for Internet Users (Doctoral dissertation, Utica College), 2017.

[16] B. L. Bayus, "Crowdsourcing new product ideas over time: An analysis of the Dell IdeaStorm community", *Management science*, vol. 59(1), pp. 226-244, 2013.

[17] R. W. Ouyang, M. Srivastava, A.Toniolo, & T. J. Norman, "Truth discovery in crowdsourced detection of spatial events", IEEE, 2016.

[18] B. Xiang, The psychological effects of participation in crowdsourcing on customer's willingness to pay and recommend a brand, 2016.

[DOI: 10.52547/jsdp.19.1.101]

[19] M. A. Rashid, K. Deo, D. Prasad, K. Singh, S. Chand, & M. Assaf, TEduChain: A platform for crowdsourcing tertiary education fund using blockchain technology. arXiv preprint arXiv:1901.06327, 2019.

[20] P. Welinder, & P. Perona, "Online crowdsourcing: rating annotators and obtaining cost-effective labels", In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops pp. 25-32, IEEE, 2010.

[21] G. M. Leung, & K. Leung, "Crowdsourcing data to mitigate epidemics", *The Lancet Digital Health*, vol. 2(4), e156-e157, 2020.

[22] A. Drutsa, V. Fedorova, D. Ustalov, O. Megorskaya, E. Zerminova, & D. Baidakova, "Crowdsourcing Practice for Efficient Data Labeling: Aggregation, Incremental Relabeling, and Pricing", *In Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, pp. 2623-2627, 2020.

[23] F. Nooralahzadeh, & L. Øvrelid, "Syntactic dependency representations in neural relation classification", arXiv preprint arXiv:1805.11461, 2018.

[24] L. Hettinger, A. Dallmann, A. Zehe, T. Niebler, & A. Hotho, "Claire at semeval-2018 task 7: Classification of relations using embeddings", In Proceedings of The 12th International Workshop on Semantic Evaluation, pp. 836-841, 2018.

[25] B. Pratap, D. Shank, O. Ositelu, & B. Galbraith, "Talla at SemEval-2018 task 7: Hybrid loss optimization for relation classification using convolutional neural networks", In Proceedings of The 12th International Workshop on Semantic Evaluation, pp. 863-867, 2018.

[26] D. Jin, F. Dernoncourt, E. Sergeeva, M. McDermott, & G. Chauhan, "MIT-MEDG at SemEval-2018 task 7: Semantic relation classification via convolution neural network", In Proceedings of The 12th International Workshop on Semantic Evaluation, pp. 798-804, 2018.

[27] M. Gluhak, M. P. di Buono, A. Akkasi, & J. Šnajder, "TakeLab at SemEval-2018 Task 7: Combining Sparse and Dense Features for Relation Classification in Scientific Texts, In Proceedings of The 12th International Workshop on Semantic Evaluation, pp. 842-847, 2018.

[28] J. Pennington, R. Socher, & C. D. Manning, "Glove: Global vectors for word representation", In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) ,pp. 1532-1543, 2014.

محمد جعفرآباد مدرک کارشناسی و کارشناسی ارشد خود را در رشته مهندسی رایانه گرایش معماری سامانه های رایانه دریافت کرده و هم‌اکنون دانشجوی مقطع دکترای دانشگاه قم در رشته مهندسی فناوری اطلاعات است. ایشان در زمینه‌های داده‌کاوی و رمزنگاری مطالعاتی داشته است.

jafarabadm@yahoo.com

روح‌الله دیانت فارغ‌التحصیل دکترای معماری سامانه‌های رایانه‌ای دانشگاه شریف است. ایشان هیأت علمی دانشگاه قم هستند و پژوهش‌هایی را در زمینه داده‌کاوی و پردازش صوت و تصویر داشته‌اند.

rouhollah.dianat@gmail.com

[DOI: 10.52547/jsdp.19.1.101]