مجله‌بین المللی سیستم های هوش محاسباتی (۲۰۲۴) ۱۷:۸
https://doi.org/10.1007/s44196-023-00375-7

مقاله‌پژوهشی

یک‌چارچوب داده کاوی رسانه های اجتماعی مبتنی بر LDA برای اقتصاد چرخشی پلاستیک

یانگیمین‌ژو[۱] · چاندراسخار کامبهامپاتی[۱] · یونگ کیانگ چنگ[۱] · نیشیکانت میشرا[۲] · نور وولندهاری[۲] · پائولین دویتز[۳]

دریافت:۲۴ فوریه ۲۰۲۳ / پذیرش: ۲۱ نوامبر ۲۰۲۳

چکیده

تولیدانبوه زباله های پلاستیکی باعث ایجاد یک بحران فوری در حوزه سلامت عمومی در سراسر جهان شده است. اگرچه سیاست های دولتی‌و نوآوری های صنعتی نیروهای محرک برای مقابله با این چالش هستند، اما تلاش برای درک نگرش های عمومی می تواند کارایی این‌فرآیند را بهبود بخشد. رسانه های اجتماعی به راه های اصلی برای کسب اطلاعات و ابراز نظرات و احساسات توسط عموم مردم تبدیل‌شده اند. این امر ما را بر آن داشت تا با استفاده از داده های رسانه های اجتماعی، برداشت ها و واکنش های رفتاری نسبت به استفاده‌از پلاستیک را کاوش کنیم. در این مقاله، ما چارچوبی برای جمع آوری و تحلیل داده ها بر اساس رسانه های جریان اصلی در بریتانیاپیشنهاد دادیم تا نظرات عمومی در مورد پلاستیک ها را به دست آوریم. یک مدل یادگیری ماشین بدون نظارت مبتنی بر تخصیص‌نهفته دیریکله (LDA) برای تجزیه و تحلیل و خوشه بندی موضوعات به منظور مقابله با عدم حاشیه نویسی محتوای داده ها به کارگرفته شده است. سپس یک روش فرهنگ لغت اضافی برای ارزیابی احساسات نظرات ارائه شد. این چارچوب همچنین ابزارهایی برای‌تجسم مدل و نتایج برای تحریک درک های عمیق ارائه می دهد. ما اثربخشی چارچوب را با استفاده از آن برای تجزیه و تحلیل سه رسانه‌اجتماعی جریان اصلی، که در آن 6 دسته موضوعی سطح اول و 13 دسته موضوعی سطح دوم از متون نظرات مربوط به پلاستیک هاشناسایی شده اند، تأیید کردیم. نتایج نشان می دهد که احساسات عمومی نسبت به محصولات پلاستیکی به طور کلی پایدار است.توزیع مکانی-زمانی احساسات هر موضوع با تعداد تکرار آن همبستگی بالایی دارد.

کلمات‌کلیدیمصورسازی مدل · تحلیل احساسات · طبقه بندی نظرات · LDA

اختصارات

ال‌دی ای	واژگان‌پردازش زبان طبیعی
ان‌ال پی	تخصیص‌نهفته دیریکله
ترکیبات‌آلی فرار (VOC)	
آران ان	حافظه‌کوتاه مدت بلند مدت
LSTM	شبکه‌عصبی بازگشتی
جی‌آر یو	واحدبازگشتی دروازه ای
بی بی سی	نمایه سازی‌معنایی پنهان شرکت
ال‌اس آی	پخش‌بریتانیا
اچ‌دی پی	فرآیندیریکله سلسله مراتبی
اس‌وی دی	تجزیه‌مقدار تکین

امقدمه

محصولات‌پلاستیکی در دهه 1950 به طور گسترده مورد استفاده قرارگرفتند. به دلیل هزینه تولید پایین، شدت متوسط، سهولت پردازش‌و راحتی استفاده، پلاستیک تقریباً در همه جا مورد استفاده‌قرار گرفته و با اهمیت روزافزون در مدیریت پسماند شناخته‌شده است. گزارش بنیاد الن مک آرتور نشان می دهد که درحال حاضر سالانه حدود 500 میلیون تن پلاستیک در جهان تولیدمی شود، اما تنها 10٪ از آنها بازیافت می شوند. علاوه بر این،بسیاری از زباله های پلاستیکی از سیستم دفع نشت می کنند وبه یک تهدید زیست محیطی تبدیل می شوند. در نتیجه، میزان پلاستیک‌در زباله های شهری از 1٪ به بیش از 10٪ افزایش یافته است.[1]این مشکل نشان می دهد که طراحی، تولید و استفاده از پلاستیک هاخطی است، نه یک الگوی دایره ای. به عنوان مثال، با درنظر گرفتن بسته بندی پلاستیکی، حدود ۳۲٪ از تولید جهانی در محیط‌طبیعی مانند اقیانوس و هوا رها شده و حدود ۴۰٪ از آن دفن شده‌است. فقط

* یانگیمین‌ژو
y.xue-2019@hull.ac.uk

چاندراسخارکامباماپاتی
c.kambhampati@hull.ac.uk

یونگ‌کیانگ چنگ
y.cheng@hull.ac.uk

۱ گروه‌علوم و فناوری کامپیوتر، دانشگاه هال، هال 7RX HU6، انگلستان

۲ دانشکده‌بازرگانی، دانشگاه هال، هال HU6 7RX، انگلستان

۳ گروه‌جغرافیا، زمین شناسی و محیط زیست، دانشگاه هال، هالHU6 7RX، انگلستان

٪۱۰از بسته بندی های پلاستیکی بازیافت شده اند و تنها ۲٪ از آنها به همان‌سطح از بازسازی دست یافته اند.۲در مقابل، در یک سیستم چرخشی،محصولات (از جمله بسته بندی) به گونه ای طراحی می شوندکه تا حد امکان دوام بیاورند، قابل استفاده مجدد باشند و در پایان‌عمر قابل بازیابی باشند.۱سیاست گذاران در بریتانیا، اتحادیه اروپا وجاهای دیگر در حال تدوین سیاست هایی برای ممنوعیت پلاستیک های‌یکبار مصرف و بهبود بازیابی پلاستیک ها هستند. ادبیات[۳،۵] دیدگاه های عمومی در مورد محصولات پلاستیکی را به عنوان‌مرجعی برای سیاست گذاران جهت تدوین سیاست های مربوطه‌و همچنین برای شرکت ها در تدوین برنامه های تجاری، خلاصه‌کرده اند.۴].

درروانشناسی، علوم شناختی و جامعه شناسی، مطالعات نشان‌داده اند که افراد از نظرات برای هدایت تصمیمات و رفتارهای‌فردی خود استفاده می کنند.۹] در مطالعه مایا بی، نشان داده‌شده است که چگونه رسانه های اجتماعی برای عموم اعتیادآورهستند [۲۸] اما برای دستیابی به یک اقتصاد چرخشی پلاستیکی‌مؤثر، باید بتوان طیف وسیعی از نظرات عمومی را درک کرد.برای مثال، تحمل عمومی برای افزایش هزینه های محصولات زیست محیطی‌چقدر است؟ مردم تا چه حد از علل و پیامدهای سوءمدیریت پلاستیک آگاه هستند و با آنها همدردی می کنند؟ از آنجایی‌که رسانه های اجتماعی به تدریج به رسانه ای مهم برای مردم‌جهت کسب و انتشار اطلاعات تبدیل می شوند، روزانه حجم عظیمی‌از داده های رسانه های اجتماعی و نظرات در اینترنت تولیدمی شود. یکی از رویکردها برای دستیابی به این بینش ها، تحلیل‌نظرات در رسانه های اجتماعی است که بر رفتار عمومی از کسب‌اطلاعات گرفته تا رفتار پس از خرید نیز تأثیر گذاشته است. [۶–۸] علاوه بر این، صدها میلیون کاربر اینترنت از طریق اینترنت و رسانه های‌اجتماعی به انواع مختلف اطلاعات مرتبط با پلاستیک دسترسی‌دارند. در مقایسه با سایر کانال های ارتباطی سنتی، رسانه های‌اجتماعی بسیار کارآمدتر هستند. به طور کلی، با افزایش‌تعداد نظرات، موضوعات و احساسات در مورد آنها به سرعت‌گسترش یافته و تکامل یافته اند و باعث نوسانات روانی و تغییرات‌عاطفی در سراسر جامعه شده اند.۱۰چگونگی حذف مؤثر نظرات‌اسپم، یافتن نظرات ارزشمند و ارائه آنها به خوانندگان، یا ارائه‌آنها برای تحلیل بیشتر افکار عمومی و وظایف متن کاوی، ارزش های‌کاربردی مهمی دارند.

امروزه،افراد کمی در مورد موضوعات خاص مشابه دیدگاه ها درمورد محصولات پلاستیکی، به تحلیل افکار عمومی می پردازند. مطالعات‌اغلب به این نتیجه می رسند که بخش خصوصی باید اقدامات‌بیشتری برای کاهش ضایعات پلاستیکی انجام دهد. به دلیل‌تفاوت در تخصص و انگیزه مفسران آنلاین، نظرات آنلاین اغلب‌از نظر محتوا و احساسات غیرقابل پیش بینی هستند. [۱۱]. این‌نظرات حاوی اطلاعات متنی می توانند موضوع و احساس نظر رابررسی کنند تا دیدگاه نظر را تحلیل کنند. هدف از این مطالعه، به دست‌آوردن دیدگاه های عمومی در مورد موضوع از بررسی های مرتبط‌با پلاستیک است تا واکنش عموم به اقتصاد پلاستیک آشکارشود. ما نظرات مبتنی بر داده های رسانه های اجتماعی را تجزیه‌و تحلیل می کنیم تا ویژگی های زمانی و مکانی و ویژگی های توزیع‌مکانی و زمانی موضوعات مختلف را با استفاده از یک مدل استخراج‌و طبقه بندی موضوع نامشخص دریافت کنیم [۱۲]. در همزمان،تعداد موضوعات ناشناخته طبقه بندی می شود و کیفیت‌نتایج طبقه بندی ارزیابی می شود.

دراین مقاله، ما چارچوبی را برای ساده سازی جنبه های مختلف‌نظرکاوی در مورد پلاستیک ها بر اساس تخصیص دیریکله پنهان(LDA) پیشنهاد می کنیم [1].۱۳]، یک روش مدل سازی موضوعی‌بدون نظارت که می تواند موضوعات بالقوه را در اسناد شناسایی‌کند، برای بررسی موضوعات پنهان یا کلماتی که اغلب در مجموعه‌ظاهر می شوند، اتخاذ شده است. این کلمات بر اساس احتمالات‌سند موضوعی و نزدیکی کلمه-موضوع محاسبه می شوند.برخلاف روش سنتی استفاده از تجربه برای قضاوت در موردتعداد و نتیجه طبقه بندی متن، این مقاله دو روش تخمین را معرفی‌می کند و نتایج را برای انتخاب نتایج طبقه بندی متن مناسب،مصورسازی می کند. پس از تعیین جنبه های مختلف هر نظر،نتایج طبقه بندی را مصورسازی می کنیم تا درک شهودی تر و واضح تری‌از فرآیند و دامنه موضوعات ایجاد کنیم. این به خلاصه کردن‌محتوای موضوع کمک می کند. برای انجام طبقه بندی احساسات،ابتدا امتیاز احساسات هر کلمه را از لیست امتیاز کلمات‌تعیین شده به دست می آوریم و سپس نتیجه امتیاز احساسات‌نظرات کامل را برای نرمال سازی محاسبه می کنیم [۱۶ درنهایت، از طریق طبقه بندی موضوعی و امتیازدهی به احساسات،می توان یافته های آموزنده ای از برداشت های عمومی‌از پلاستیک های بازیافتی و تغییرات رفتاری را برای کمک به‌سیاست گذاری ارائه داد.

ادامه‌مقاله به شرح زیر سازماندهی شده است: بخش.۲بخش مربوطه،اثر مرتبط است.۳جزئیات چارچوب را در بخش ... ارائه می دهد.۴، نتایج مختلف را ارائه می دهیم و نتایج خود را با سایر روش های‌پیشرفته مقایسه می کنیم. مقاله را در بخش نتیجه گیری می کنیم.۵.

۲کار مرتبط

دراینجا، ما کارهای مرتبط مورد استفاده در این تحقیق، به طور خاص،پیشینه NLP، LDA، ارزیابی مدل موضوعی و تحلیل احساسات‌را ارائه می دهیم.

۲.۱پردازش زبان طبیعی

پردازش‌زبان طبیعی (NLP) پلی بین زبان ماشین و زبان انسان است.این رشته به مطالعه چگونگی فهم زبان انسان توسط کامپیوترهامی پردازد. [۱۷]. این فناوری اغلب بالاترین وظیفه و قلمروهوش مصنوعی را در بر می گیرد. این شاخه ای از هوش مصنوعی‌است و نقطه تلاقی علوم کامپیوتر و زبان شناسی است. NLP همچنین به عنوان زبان شناسی محاسباتی در رشته زبان شناسی‌شناخته می شود. تحقیقات اولیه پردازش زبان طبیعی‌شامل تحلیل دستوری، تحلیل معنایی و درک متن است. هرتحلیل به یک مرحله پیش پردازش نیاز دارد که شامل برخی از فناوری های‌مهم مانند توکن سازی است. در این فرآیند، متن ورودی‌از سند به کوچکترین واحدها (کلمات، عبارات، شکلک ها و غیره) تقسیم می شود. در بیشتر موارد، این واحد در

شکل‌کلمات یا جملات یا پاراگراف ها. فرآیند نرمال سازی می توانندتمام حروف را به حروف کوچک تبدیل کند. علائم نگارشی‌و ریشه یابی را حذف کنید، که یک متن کامل را به کلمات اساسی‌تبدیل می کند [۱۸]مرحله بعدی تولید یک مدل، یک فرهنگ لغت‌و یک پیکره زبانی برای فیلتر کردن کلمات با فرکانس پایین و ساخت‌بردارها است.

۲.۲فاصله نزدیک (LDA)

[LDA۱۳] یک مدل احتمالی مولد است که می تواند برای طبقه بندی‌پیکره ها استفاده شود. LDA یک مدل احتمال بیزی سلسله‌مراتبی سه سطحی است که شامل کلمه، موضوع و سند می شود.اصل این روش این است که فرض کنیم هر کلمه از یک موضوع‌بالقوه پنهان در پشت آن استخراج می شود. در طول فرآیندمولد، انتخاب موضوع و انتخاب کلمه دو فرآیند تصادفی هستند.در فرآیند تولید، انتخاب موضوع و انتخاب کلمه دو فرآیند تصادفی‌هستند. ابتدا، برای هر سند، یک موضوع را از توزیع موضوع‌استخراج کنید. سپس، یک کلمه از توزیع کلمه مربوط به موضوع‌انتخاب شده استخراج می شود. فرآیند فوق را تا زمانی که هرکلمه در سند پیمایش شود، تکرار کنید. فرض می شود که هم سندی‌که متعلق به یک موضوع است و هم موضوعی که می تواند توسط‌یک کلمه نمایش داده شود، از یک توزیع چندجمله ای پیروی می کنند.به طور خلاصه، هدف LDA شناسایی موضوعات از اسناد است،یعنی تبدیل ماتریس سند-کلمه به یک ماتریس سند-موضوع(توزیع) و یک ماتریس موضوع-کلمه (توزیع).

مزیتLDA این است که یک یادگیری ماشین بدون نظارت است و به‌مجموعه داده های آموزشی که به صورت دستی حاشیه نویسی شده اند،متکی نیست، یعنی تنها ورودی ها مجموعه اسناد و تعداد موضوعات‌هستند. علاوه بر این، LDA همیشه می تواند کلمات نماینده ای‌برای توصیف هر موضوع پیدا کند.

برای‌مدل متن پنهان LDA، عیب اصلی این است که موقعیت مجموعه ای‌از کلمات را در متن در نظر نگرفته است. از این رو، نمی تواندمعانی مختلفی را که کلمات یکسان با ترتیب های مختلف می توانندبیان کنند، تشخیص دهد. علاوه بر این، از آنجا که متون طولانی‌حاوی کلمات بیشتری هستند، تطبیق موضوعات آنها دشوار است.مشکل دیگری که باید حل شود این است که کلماتی که از موضوعات‌مختلف در LDA تشکیل شده اند، دوباره استفاده می شوند.این امر منجر به همپوشانی موضوعات به جای استقلال آنها می شود.

اول،تعریف می کنیم‌دی‌همانطور که سند تنظیم شده است.تی مجموعه ای‌از موضوعات است. در مجموعه سنددی، سند می توانندبه عنوان یک دنباله کلمه در نظر گرفته شود <۱،۲،....،ن>،د شاملn کلمه است. تمام کلمات مختلف در این‌دی‌در یک مجموعه‌بزرگ واژگان (VOC) ترکیب می شوند. در D، احتمال هر سند مربوط به موضوعات مختلف‌د پی‌تی۱،پی تی۲،...،پی تی‌ن> ، کجا پی‌تی‌من‌احتمال این است که d متناظر بامن‌موضوع هفتم در تی نگاه‌شهودی به محاسبه،پی تی‌من=تی‌من‌ان، کجاتی‌من‌تعداد کلمات مربوط‌به‌من‌موضوع هفتم درد، ون‌در دتعداد کل کلمات است. برای‌هرتی‌درتی‌احتمال تولید کلمات متفاوت‌تی< پی دبلیو۱،پی دبلیو ۲...،پی دبلیون> کجاپی دبلیو‌من‌احتمال آن است که‌تی‌تولید می کند من‌کلمه هفتم در ،روش‌محاسبه نیز بسیار شهودی است .VOCپی دبلیو‌من= دبلیو‌من/ ن، کجادبلیو‌من‌تعداد است‌من‌کلمه ام در VOC مربوط به موضوع، وتعداد کل کلمات مربوط به موضوع‌ن.

ص=)d | w(پ×)t | w(پ(زمان | زمان) (1)

این‌فرمول، فرآیند اصلی است که لایه موضوع را به عنوان لایه میانی‌که کلمات و سند را به هم متصل می کند، در نظر می گیرد و سپس‌احتمال ... را می دهد.دبلیودر سنداز طریق دوتی‌با استفاده از جریان‌دوتی، می توانیم محاسبه کنیم ص(دبلیو | د)یک کلمه در یک سند را وقتی که با هر موضوعی مطابقت دارد، بررسی‌کنید و سپس موضوع مربوط به کلمه را بر اساس این نتایج به روزرسانی‌کنید. سپس، اگر به روزرسانی، موضوع مربوط به کلمه راتغییر دهد، به نوبه خود تأثیر خواهد گذاشت.دو تی.

درزیر فرآیند یادگیری الگوریتم LDA آمده است: در ابتدا،دوتی‌به طورتصادفی اختصاص داده می شوند (برای همه دوتی‌سپس، فرآیناصلی فوق را به طور مداوم تکرار کنید، تا نتیجه همگرایی نهایی،که خروجی LDA است.

برای‌من‌کلمه هفتم‌من‌در یکی از موضوعات سند ها، هنگامی که مربوط‌به کلمه، فرمول می تواند به صورت زیر جی، سپس موارد فوق اصلاح‌شود:

( ) ( ) ( )
پ‌جی‌دبلیو‌من | دها =پ دبلیو‌من | تی‌جی *پ ت‌جی | دها. (2)

هنگام‌برشمردن موضوعات درتی‌برای بدست آوردن همه‌پ‌جی(که‌پ(و) ، آن‌مقدارجی‌است‌ا~ک‌سپس، طبق این مقادیر احتمال، برابر است بامن کلمه‌هفتم‌من‌درهایک موضوع انتخاب کنید. ساده ترین ایده این است که‌من که‌به حداکثر می رساندپ‌جی(من | ها) (توجه داشته باشید که فقطجی‌یک متغیردر این فرمول است)، که argmax است.

[ج]پ‌جی(من | ها).

درمرحله بعد، اگرمن‌کلمه اصلی، من‌درهاموضوعی را از بین موضوعات انتخاب می کند آن‌را تحت تأثیر قرار خواهد داد دوتی(می توان از آن استنباط کرد فرمول‌محاسبه دو بردار فوق). این تأثیر همچنین به محاسبه ... منتقل‌خواهد شد.پ(دبلیو | د) بالا. محاسبه کنیدپ(دبلیو | د)برای همه‌در همه حال‌دردی‌و موضوع را به عنوان یک موضوع دوباره انتخاب‌کنید. این روش را تا بعد از آن اجرا کنیدن‌تکرارهای حلقه، نتیجه‌نهایی مورد نیاز LDA را پوشش داده و دریافت می کند.

دراینجا، ما دو احتمال را به دست آوردیم‌پ(کلمه | موضوع)و پ (موضوع | سند).

۲.۳ارزیابی مدل موضوعی

اکثرروش های سنتی از بازرسی بصری یا دانش قبلی برای ارزیابی عملکردمدل های انتخاب موضوع استفاده می کنند. شهودی ترین روش،قضاوت دستی در مورد موضوع استخراج شده است، اما بدیهی‌است که این کار زمان بر است. قضاوت بصری دستی عمدتاً شامل ارزیابی نتایج خوشه بندی با استفاده از فناوری تجسم ومعرفی معنای موضوعات توسط کلمات کلیدی تولید شده از مدل‌موضوع است [۱۵].

ازسوی دیگر، روش های ارزیابی خودکار شامل ارزیابی اثر خوشه بندی‌توسط ضریب سیلوئت و انسجام برای اندازه گیری اینکه‌آیا کلمات موجود در موضوع ... هستند یا خیر، می شوند.

ضریب‌سیلوئت (Silhouette Coefficient) یک روش ارزیابی عملکردخوشه بندی است که اولین بار توسط پیتر جی. روسو در سال ۱۹۸۶پیشنهاد شد. این روش ترکیبی از انسجام (cohesion) و تفکیک پذیری(resolution) است. این روش یکی از روش های ارزیابی‌تأثیر الگوریتم های مختلف یا حالت های عملیاتی مختلف بر نتایج‌خوشه بندی بر اساس داده های اصلی یکسان است. برای هر نمونه*من* در داده ها، میانگین فاصله از*من*به نمونه های دیگر در این خوشه‌است*الف(ا)*حداقل میانگین فاصله از I تا تمام نمونه های خوشه های‌دیگر، b است.*(من)*فرمول می تواند به صورت زیر باشد:

بازنویسی‌شده به صورت:

(من) =ب(من) −الف(من)\حداکثر {الف(من) ،ب(من)} (3)

(من) = ۱ −الف(من)\ب(من) ،الف(من)<ب(من) ؛ ۰، الف(من) =ب(من) ؛ ب(من)\الف(من) −۱،الف(من)>ب(من) (4)

(1) میانگین فاصله را محاسبه کنید*الف(i)*بین نمونه i و سایر نمونه های‌همان خوشه. چه زمانی*الف(i)*کوچکتر باشد، اثر گروه‌بندی خوشه*من*بهتر است. *(من)*به عنوان عدم تشابه درون‌خوشه ای نمونه در نظر گرفته می شود*من*.
(2) میانگین فاصله بین نمونه I و سایر نمونه ها در خوشه دیگر b برابراست با *(من)*. این به عنوان عدم تشابه بین خوشه ای نمونه‌تعریف می شود.*من*:

من=دقیقه‌ب من۱،ب من۲،...ب آی کی

(3) چه زمانی من نزدیک به ۱ باشد، خوشه بندی نمونه*من*منطقی است.اگر من نزدیک به -1 باشد، عکس آن صادق است.

انسجام‌موضوعی به این صورت است که با مقایسه شباهت معنایی‌بین کلمات با امتیاز بالاتر، به موضوع امتیاز داده می شود. درموضوع.

(هسته V من،وی جی،) = ورود به سیستم پی وی من،وی جی+ ((...))) − لاگ پی وی من پی وی جی (( ) ( ))) (5)

*پنجم*گروهی از کلمات است که برای توصیف موضوع استفاده می شود،*ε* به معنای اطمینان از این است که نمره یک عدد حقیقی‌را برمی گرداند. در اینجا، هر چه اپسیلون کوچکتر باشد، نتیجه‌کوچکتر خواهد بود. احتمال کلمه*ص*با شمارش فراوانی کلمات‌روی مجموعه داده محاسبه می شود. این الگوریتم بر اساس‌مجموعه داده اصلی مدل موضوع آموزشی است و به مجموعه‌داده خارجی متکی نیست.

۲.۴تحلیل احساسات

تحلیل‌احساسات فرآیند تحلیل، پردازش، القا و استدلال ذهنی متن‌با رنگ های احساسی است [۱۹،۲۰]تحلیل احساسات می تواند ازروش های سنتی مبتنی بر فرهنگ لغت احساسات یا روش های مبتنی‌بر یادگیری عمیق استفاده کند.۲۵روش مبتنی بر فرهنگ لغت‌عمدتاً شامل تدوین مجموعه ای از فرهنگ های احساسات و قوانین،تجزیه جملات، تجزیه و تحلیل و تطبیق است.

فرهنگ‌لغت های متن (به طور کلی تحلیل اجزای کلام، تحلیل وابستگی‌نحوی)، محاسبه مقادیر احساسی، و در نهایت استفاده از مقادیراحساسی به عنوان گرایش احساسی متن. مبنای قضاوت [۱۶،۲۱،۲۲]. بر اساس طبقه بندی احساسات مبتنی بر یادگیری عمیق،جمله ابتدا بر اساس طبقه بندی احساسات مبتنی بر یادگیری‌عمیق است. ابتدا جمله پیش پردازش می شود، مانند تقسیم بندی‌کلمات، کلمات توقف و تبدیل ساده شده و سنتی. سپس،کدگذاری بردار کلمات و همچنین استخراج ویژگی با استفاده‌از RNN (شبکه عصبی بازگشتی) مانند LSTM (حافظه کوتاه مدت‌بلند) یا GRU (واحد بازگشتی دروازه دار) انجام می شود.

[۲۳مراحل عملیاتی عبارتند از: سند باید از جملات و برچسب ها تشکیل‌شده باشد. جمله را توکنیزه کنید،
بنابراین‌لیستی از کلمات آن را نشان می دهد. سپس، از طریق ویژگی های ساده‌کلمات یونیگرام، به ترتیب از نمونه های ذهنی و عینی برای حفظ توزیع‌متعادل و یکنواخت کلاس در مجموعه آموزش و مجموعه تست استفاده‌می شود. از این ویژگی ها برای به دست آوردن یک نمایش مقدار ویژگی‌از مجموعه داده خود استفاده می شود. سپس باید
طبقه بندی کننده‌را روی مجموعه آموزش آموزش دهیم و در نهایت نتایج راخروجی دهیم.

۳چارچوب پیشنهادی

این‌چارچوب شامل چهار بخش است: جمع آوری و پیش پردازش داده ها،طبقه بندی موضوعات، انتخاب مدل و مصورسازی، و تحلیل احساسات،همانطور که در شکل نشان داده شده است.۱.

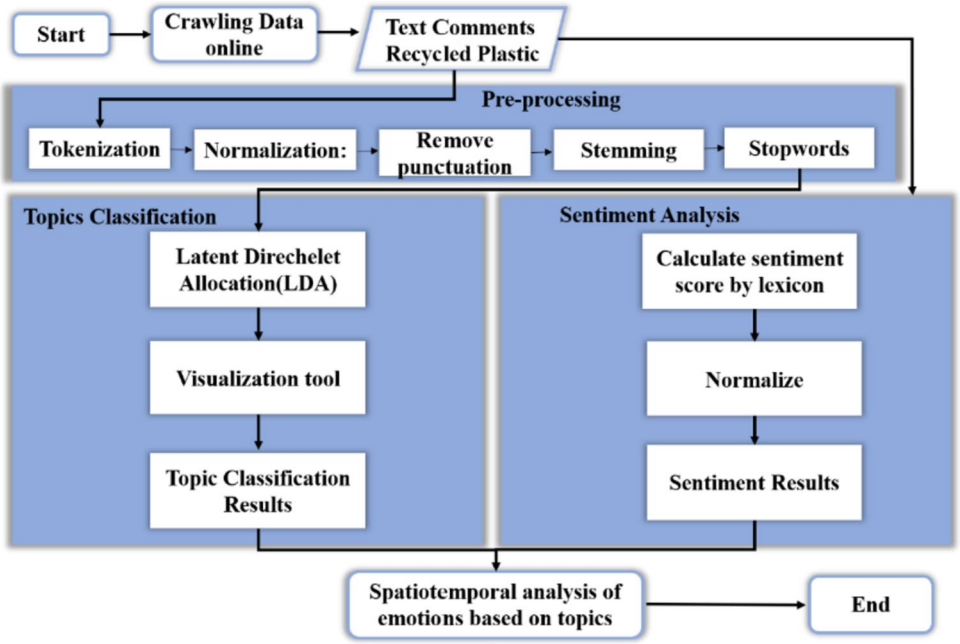
درمرحله اول، داده های نظرات را می توان از رسانه های اجتماعی‌آنلاین بررسی کرد. در مرحله بعد، نظرات پیش پردازش می شوندتا نویز محتوای تکراری یا نظرات با ساختار غیرمتنی برای پردازش‌در نظرات فیلتر شوند. پس از پیش پردازش، یادگیری ماشین‌بدون نظارت LDA بر روی مجموعه داده های آموزشی آموزش‌داده شد. پس از آن، یک ارزیابی انجام شد.
فرآیند و یک ابزار تجسم برای کمک به ما در انتخاب تعداد موضوعات وخلاصه کردن موضوعات. در عین حال، مدل ارزیابی احساسات به‌هر نظر یک امتیاز احساسی می دهد. این امتیاز می تواند برای طبقه بندی‌احساسات یا مطالعه تغییرات در احساسات استفاده شود.در نهایت، با ترکیب طبقه بندی موضوعی و نمرات احساسی،می توانیم یک سری زمانی از تغییرات احساسی در جنبه های‌مختلف محصولات پلاستیکی به دست آوریم.

ازآنجایی که داده های مورد بررسی از نظرات خوانندگان رسانه های‌اصلی بریتانیا از جمله بی بی سی، گاردین و میل آنلاین درزیر اخبار مربوط به پلاستیک قابل بازیافت گرفته شده است، داده هادارای ویژگی های غیرقابل پیش بینی بودن و بدون برچسب های‌طبقه بندی موجود هستند. بنابراین، ما باید از تکنیک های‌استفاده کنیم که نیازی به مجموعه آموزشی برای طبقه بندی‌نظرات ندارند. در اینجا، محقق یک روش یادگیری بدون نظارت‌مبتنی بر کلمه به نام LDA را معرفی می کند.

۳.۱جمع آوری و پیش پردازش داده ها

چندرسانه مانند بنگاه سخن پراکنی بریتانیا (بی بی سی)، گاردین و میل‌آنلاین تا حد زیادی

شکل۱روش تحقیق



مصرف‌اخبار آنلاین در بریتانیا را تحت سلطه خود قرار داده است [۲۴] اعتمادخوانندگان به این رسانه ها به این معنی است که صرف نظر از اینکه‌خبر چیست، آنها مستقیماً و مکرراً جستجو و مطالعه می کنند. نظرات‌بیان شده پس از خواندن خبر اغلب ارتباط زیادی با موضوع خبردارد.

این‌مرحله، فرآیندی است که توسط الگوریتم ۱ انجام می شود، ازجمله: متن نظرات اصلی حاوی اطلاعات تداخلی مانند فاصله ها،لینک های http و علائم نگارشی است. برای حذف نویزو بهبود کارایی تقسیم بندی کلمات، داده های اصلی باید فیلتر متن‌شوند. ما از عبارات منظم پایتون برای فیلتر کردن متن اصلی رسانه های‌اجتماعی و حذف اطلاعات تداخلی (مانند لینک های http، علائم نگارشی)، کلمات بی کیفیت، متن بی کیفیت و متن تکراری‌استفاده می کنیم. شکل۲

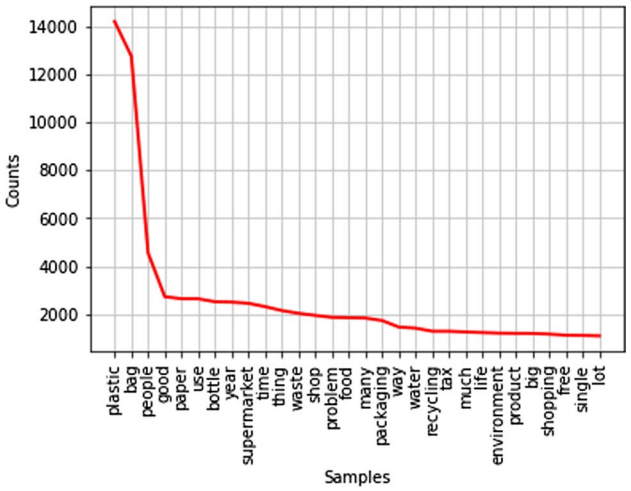
الگوریتم۱

```
Start
1.  Take the comments that have been crawled in the previous stage.
    For all comment in dataset:
        remove_unwanted_word
        remove_short_word
        remove_stop_word_
        make_lowercase
        filter_noun_adjective
        return tokenized_reviews
2.  Calculate frequency words in the data.
End
```

رتبه بندی30 کلمه ای که بیشترین تکرار را در داده های جمع آوری شده ما داشته اند رانشان می دهد.

۳.۲طبقه بندی موضوعات

چارچوب‌استخراج و طبقه بندی موضوع بر اساس مدل موضوعی LDA ساخته شده است و احساسات عمومی در مورد موضوع به صورت‌لایه ای از متون مرتبط در رسانه های اجتماعی به دست می آید.از مدل موضوعی LDA برای استخراج موضوع استفاده می شودتا توزیع احتمال موضوع هر متن و توزیع احتمال کلمه هر موضوع‌تولید شود. پس از آن، داده های نمونه از موضوع حاشیه نویسی‌شده در کل مجموعه داده ها طبقه بندی می شوند.



شکل۲ 30کلمه پرتکرار در داده ها

مدل هایLDA قادر به تشخیص موضوعات در اسناد و کاوش در پیکره‌برای اطلاعات پنهان هستند و طیف گسترده ای از کاربردها را در سناریوهایی‌مانند تجمیع موضوعات، استخراج اطلاعات از متن بدون ساختارو انتخاب ویژگی دارند. این مدل می تواند (1) الگوهای موضوعی‌پنهان در پیکره را کشف کند؛ (2) اسناد را بر اساس موضوعات‌حاشیه نویسی کند؛ و (3) از حاشیه نویسی برای سازماندهی،جمع آوری، خلاصه سازی و بازیابی اسناد استفاده کند. LDA یک مدل احتمالاتی برای حل مسئله مدل سازی موضوع متن است.این یک روش یادگیری بدون نظارت است که به طور خودکار الگوهای‌پنهان را کشف می کند.

ساختارهای‌موضوعی را از تعداد زیادی سند استخراج می کند و هر سندرا به یک یا چند موضوع اختصاص می دهد (شکل 1).۳) [۲۵].

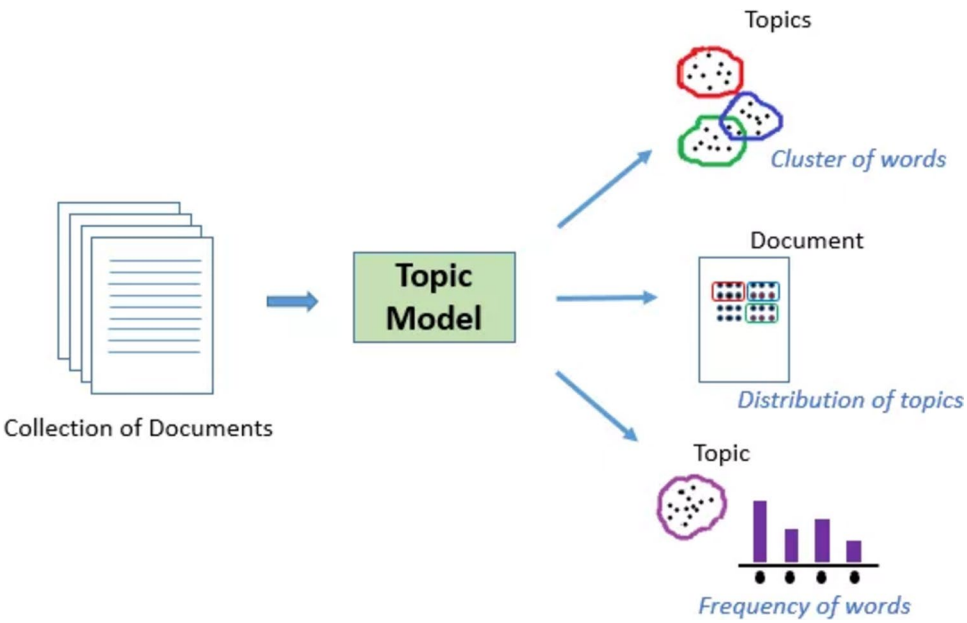
هنگام‌ساخت یک مدل موضوعی، برخی پارامترهای اصلی باید ازقبل تنظیم شوند. این پارامترها شامل تعداد موضوعات K، آلفای‌پیشینی توزیع موضوع، بتای پیشینی توزیع کلمات و تعداد اسنادی‌است که باید در هر اندازه قطعه بلوک آموزشی استفاده شوند.تعداد کل قبولی های ارزیابی آموزشی. در این آزمایش حفظ شده‌است. ک=۵۰ / ک=۰.۱).

چگونه‌تعداد مناسب موضوعات را انتخاب کنیم؟ هرچه موضوعات‌بیشتری انتخاب کنیم، موضوعات خاص تر خواهند بود.با این حال، انتخاب موضوعات زیاد ممکن است تشخیص نظرات‌را به روشی معنادار غیرممکن کند. در عین حال، تعداد کم موضوعات‌منجر به ترکیب نظرات به سمت جنبه های یکسانی می شودکه باید به دسته های مختلفی تعلق داشته باشند.

بامقایسه کیفیت موضوعات تولید شده، تعداد بهینه موضوعات‌انتخاب می شود. مرحله بعدی محاسبه و مقایسه احتمالات‌موضوعاتی است که یک بررسی واحد به آنها تعلق دارد وتعیین موضوعات نظر. در نهایت، از این موضوع برای طبقه بندی‌همه اسناد استفاده می شود.

مباحث‌مختلفی از مدل LDA ساخته شده اند [۲۶]، که در آن هر موضوع‌ترکیبی از کلمات کلیدی است و هر کلمه کلیدی وزن خاصی‌به موضوع می دهد. از لیست کلمات کلیدی، کلماتی که به درک‌و خلاصه سازی موضوعات کمک می کنند. این کلمات کلیدی ووزن ها برای خلاصه سازی محتوای موضوع استفاده خواهند شد.

برای‌هر فرآیند LDA، مراحل زیر به عنوان الگوریتم ۲ دنبال می شوند:



شکل۳یک گردش کار از مدل قالب‌معمولی [۲۵]

الگوریتم۲

Start
1. From the previous pre-processed step, get the term list.
2. Selected LDA model, calculate the probability of each word appearing in the document and corpus, also the probability of proximity of document-topic and word-topic.
for all topics $k \in [1, K]$:
sample mixture components $\vec{\varphi}k \sim \mathrm{Dir}(\vec{\beta})$
end for
for all comments $m \in [1, M]$:
sample mixture proportion $\vec{\gamma}m \sim \mathrm{Dir}(\vec{\alpha})$
sample document length $N_m \sim \mathrm{Poiss}(\xi)$
for all words $n \in [1, N_m]$:
sample topic index $z_{m,n} \sim \mathrm{Mult}(\vec{\gamma}m)$
sample term for word $w_{m,n} \sim \mathrm{Mult}(\vec{\varphi}z_{m,n})$
end for
end for
Parameters and variables:
K: the number of topics
$\vec{\varphi}k$: word distribution for topic k
$\vec{\alpha}$, $\vec{\beta}$: Dirichlet parameters
M: the total number of comments
V: vocabulary size
$\vec{\gamma}m$: topic distribution for comments m
N_m : the length of comments m
$z_{m,n}$: topic index of nth word in comments m
$w_{m,n}$: a particular word for word placeholder (m, n)
3. Display the results of topic classification (key words and word frequency under Hidden topic content)
4. Summary topics.
End

۳.۳انتخاب مدل و تجسم آن

درمرحله بعد، ما از یک ابزار تعاملی مبتنی بر وب برای نمایش نتایجمدل LDA استفاده می کنیم [۱۴]، از جمله معنی هر موضوع، شیوعهر موضوع و ارتباط هر موضوع.

.یکسیستم تجسم تعاملی مبتنی بر وب است. این سیستم نه تنهاترتیب جهانی کلمات را در یک موضوع ارائه می دهد، بلکه توزیعموضوعی کلمات خاص و کلمات انحصاری را نیز نمایش می دهد.به نوبه خود، یک معیار جدید برای ارتباط پیشنهاد می کند.این مقاله همچنین از نتایج یک مطالعه کاربری استفاده می کندتا نشان دهد که انتخاب کلمات به ترتیب نزولی احتمال برایتفسیر موضوع بهینه نیست. برای خلاصه سازی موضوع، بهینه ترینتوضیح صرفاً کلمات با بیشترین فراوانی یا کلمات کاملاً خاصنیست. بلکه باید از کلماتی که مرتبط تر هستند برای توضیح طبقه بندیموضوع استفاده شود LDAvis

رابطکاربری عملیات در شکل نشان داده شده است.۴می توان بانگه داشتن ماوس روی دایره سمت چپ، موضوع خاص را مشاهدهکرد. پس از انتخاب، سمت راست 30 کلمه برتر مرتبط با اینموضوع را نشان می دهد. می توان از این کلمات برای خلاصه کردنمیانگین این موضوع استفاده کرد. اندازه دایره، فراوانی این موضوعرا نشان می دهد. در اینجا، ما از تحلیل چندبعدی استفاده می کنیمو مؤلفه های اصلی را استخراج می کنیم.

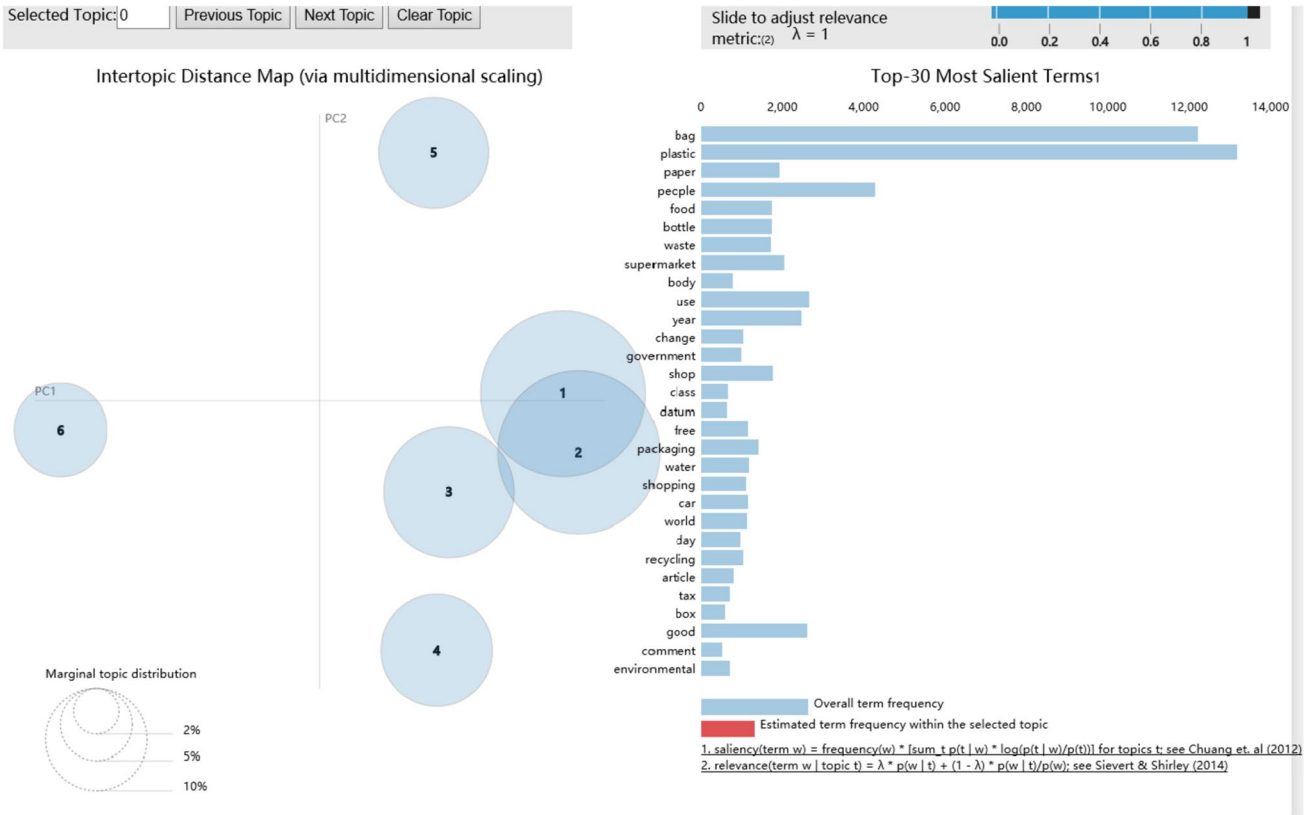
به‌عنوان ابعاد، و توزیع موضوع را روی این دو بعد قرار دهید. فاصله‌بین موضوع نشان دهنده ارتباط آنها است. ترتیب کلمات درسمت راست، وزن کلماتی را که در موضوع نقش دارند نشان می دهد.ارتباط را می توان با پارامتر تنظیم کرد$\lambda \subseteq [0,1]$. تغییر مقدارλترتیب وزن کلمات در موضوع را تغییر می دهد، که به تجسم‌محتوای اصلی موضوع کمک می کند. هرچه بزرگتر باشدλ هرچه‌تعداد کلمات بیشتر باشد، فرکانس آنها کمتر است.λ، کلمات‌خاص تر.

$$\text{ارتفاع} = p(w|t)*\lambda + (1-\lambda)*\frac{p(w|t)}{p(w)}. \quad (6)$$

هدف‌از طبقه بندی موضوعات، داشتن حداقل تعداد ممکن از دسته بندی های‌موضوعی و نسبتاً مستقل است. زیرا تعداد زیاد دسته بندی های‌موضوعی بی معنی است. همپوشانی موضوعات، نتایج‌تحقیق را گیج کننده خواهد کرد. این چارچوب دارای توابع انعطاف پذیری‌برای تنظیم تعداد دسته بندی ها و موضوعات است.

۳.۴تحلیل احساسات

مااز ابزار تحلیل Vader متعلق به NLTK استفاده می کنیم [۱۶در این چارچوب‌به عنوان الگوریتم ۳. VADER (فرهنگ لغت و استدلال احساسات‌آگاه از ظرفیت)، یک ابزار تحلیل احساسات مبتنی بر فرهنگ‌لغت و قانون است که نیازی به آموزش یا ... ندارد.



شکل۴تجسم مدل موضوعی

الگوریتم۳

```
Start
1. Take the original comments.
2. By iterating the SentiText lexicon, then calculate whole sentiment score for comment.
   for w in lexicon:
      w_score = w_dict
   end for
   tol_score =sum(w_score)
3. Normalize the score.
   fin_score = total_score \ sqrt (total_score * total_score + alpha)
4. Save the result.
End
```

به‌منظور کسب امتیاز ارزیابی احساسات، سفارشی سازی شده است و به طور ویژه بااحساسات بیان شده در رسانه های اجتماعی تنظیم شده است. این ابزار برای محتوای‌رسانه های اجتماعی طراحی شده است، بنابراین در محتوایی که می توانید دررسانه های اجتماعی پیدا کنید، بهترین عملکرد را دارد.

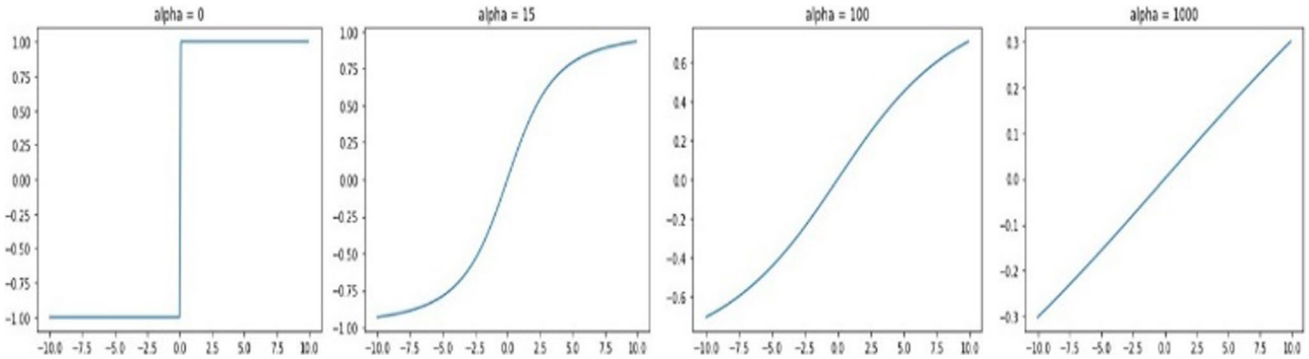
این‌ابزار برای هر نظر یک امتیاز ترکیبی ارائه می دهد. امتیاز ترکیبی‌با جمع کردن امتیازهای متناظر هر کلمه در فرهنگ لغت، سپس‌تنظیم آنها طبق قوانین و در نهایت نرمال سازی آنها بین -1 و+1 محاسبه می شود. این فرهنگ لغت شامل بیش از 9000 کلمه‌است که به صورت دستی با امتیازهای احساسی مشخص شده اند.فرمول نرمال سازی به صورت زیر است:

$$\text{کورای‌هنجار} = \frac{\text{امتیاز}}{\sqrt{\text{امتیاز}^2 + \alpha}} \quad (7)$$

درفرمول، پارامتری است که برای تنظیم توزیع فاصله ای نتایج نرمال شده‌استفاده می شود. تأثیر معادله (V) در شکل نشان داده شده‌است.۵برای طبقه بندی راحت تر احساسات، در مقاله ما آلفا = ۱۵ انتخاب شده است. این امر باعث می شود نتیجه نرمال شده درمحدوده مورد نظر یکنواخت تر باشد.

درنهایت، یک آستانه استاندارد برای طبقه بندی جملات تعیین کنید:

احساسات‌مثبت: امتیاز ترکیبی ≥ 0.05.



شکل۵چگونه آلفا بر نتیجه نرمال شده تأثیر می گذارد

احساس‌خنثی: (امتیاز ترکیبی > -0.05) و (امتیاز ترکیبی <0.05)
.
احساسات‌منفی: امتیاز ترکیبی ≤ - 0.05.

۴نتایج و تحلیل

۴.۱مجموعه داده ها

داده هااز فید خوانندگان مقالات خبری از بی بی سی، گاردین و دیلی‌میل آنلاین گرفته شده اند. به طور خاص، ما ۷۹۲۴ نظر از ۲۶ خبراز بی بی سی، با بازه زمانی از آگوست ۲۰۱۷ تا اکتبر ۲۰۱۹؛ ۲۲۸۵۷نظر از ۹۷ خبر از گاردین، با بازه زمانی از جولای ۲۰۱۶ تا اکتبر ۲۰۱۹؛و ۳۲۲۱ نظر از ۲۲ خبر از دیلی میل آنلاین، با بازه زمانی از ژانویه۲۰۱۸ تا اکتبر ۲۰۱۹ داریم.

۴.۲دسته بندی موضوعات

میزانتایج دسته بندی نظرات بر اساس موضوع را نشان می دهد. مانظرات را به شش دسته تقسیم کردیم:*محصول پلاستیکی*،*خرید*، *سیاست*،*خانواده*،*غذا* ودیگران*مردم به محصولات پلاستیکی، خرید و سیاست بیشتر از محتوای‌خانواده، غذا و غیره توجه می کنند، به ترتیب ۱۰، ۱۶۸، ۹۲۸۱ و۵۷۰۰. ترتیب کلمات کلیدی به ترتیب بزرگی ضرایب است. ضریب،فراوانی وقوع کلمه است. کلماتی که با رنگ قرمز مشخص شده اند،کلمات کلیدی هستند که برای تسهیل درک خواننده از موضوع‌خلاصه شده اند.

برای‌درک جزئیات دقیق، ما یک طبقه بندی ثانویه برای سه موضوع‌برتر با بیشترین تعداد پیام، یعنی محصولات پلاستیکی، خریدو سیاست گذاری، به صورت جدول انجام دادیم.۲.

موضوع‌محصولات پلاستیکی شامل هفت زیرموضوع است: استفاده‌مجدد، منابع، بازیافت، بسته بندی آب، آلودگی و موارد دیگر. سه‌زیرموضوع در موضوع خرید وجود دارد: بسته بندی قابل استفاده مجدد،بسته بندی پولی و بسته بندی تحویل. علاوه بر این، کسب و کار،رسانه و محیط زیست زیرموضوعات موضوع سیاست گذاری هستند.

۴.۳انتخاب مدل

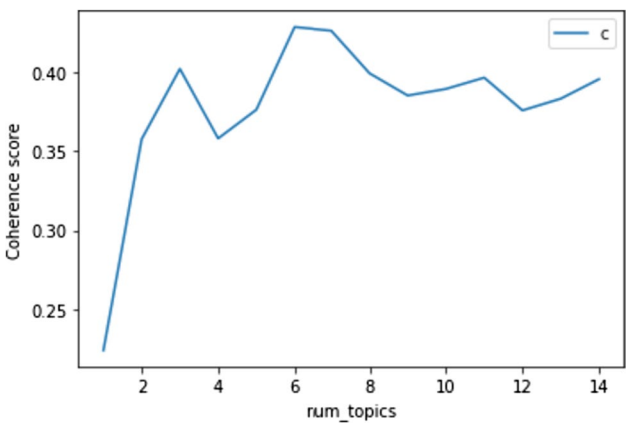
مانمرات انسجام شماره های مختلف موضوع را در شکل 1 مقایسه‌کرده ایم.۶می توانیم ببینیم که وقتی تعداد موضوعات ۶ است،امتیاز انسجام بالاترین است. جدول۳

جدول۱شرح مباحث سطح اول

	کلمات‌کلیدی	اعداد
مبحث‌اول: محصول	0.090*"پلاستیک" +0.020*"بطری" +0.020*"ضایعات" +0.013*"آب" +0.012*"روغن" +0.012*"محصول" +0.012*"بازیافت" چسبیدن" +0.011*"استفاده" +0.010*"بسته بندی" +0.010*"شیشه"	۱۰,۱۶۸
موضوع۲: خرید	'0.145*"کیسه" +0.064*"پلاستیک" +0.023*"کاغذ" +0.021*"سوپرمارکت" +0.021*"استفاده" +0.018*"فروشگاه" +0.014*"افراد- لطفا" +0.014*"رایگان" +0.013*"خرید" +0.011*"سال"	۹۲۸۱
موضوع۳: سیاست	۰.۰۳۴*"مردم" +۰.۰۲۰*"خوب" +۰.۰۱۹*"تغییر" +۰.۰۱۸*"دولت" +۰.۰۱۷*"چیز" +۰.۰۱۵*"جهان" +۰.۰۱۳*"مشکل" "مالیات" +0.013*"محیط زیست" +0.011*"کشور"*0.013+ "lem	۵۷۰۰
موضوع۴: خانواده	نفر" +0.022*"سال" +0.018*"روز" +0.016*"زمان" +0.015*"ماشین" +0.014*"مرد" +0.013*"بچه" +0.012*"فرزند" +0.011*"کفش"*0.022 هزینه" +0.010*"بسیاری"	۳۵۵۰
موضوع۵: غذا	0.027*"غذا" +0.016*"جعبه" +0.014*"بسته- "میوه" +0.012*"گوشت" +0.011*"چیز" +0.011*"خوب" +0.010*"محلی" +0.009*"بی کیفیت" +0.009*"کوچک"*0.014+ "ing"	۲۷۶۳عدد
موضوع:سایر	0.028*"body" +0.024*"class" +0.023*"datum" +0.019*"comment" +0.016*"article" +0.010*"demand" +0.009*"under- خط" +0.009*"متشکرم" +0.009*"درسته" +0.008*"سوال"	۲۱۱۹

جدول۲شرح موضوع در

موضوع‌سطح مشت	مبحث‌سطح دوم	وزن هاو کلمات کلیدی
محصولات‌پلاستیکی یوسی تی	دیگر	0.035*"داده" +0.025*"بدنه" +0.028*"لینک" +0.032*"نام" +0.033*"com" +0.024*"پترو" +0.010*-کلاس" "زیر خط" +0.009*"بزرگ" +0.007*"باد" +0.009*"leum"
	استفاده‌مجدد	0.105*"بطری" +0.052*"لیوان" +0.024*"پلاستیک" +0.020*"شیر" +0.019*"با "فروشگاه" +0.018*"سپرده" +0.015*"فروشگاه" +0.012*"افراد" +0.011*"طرح" +0.009*"سوپرمارکت"
	منبع	0.083*"پلاستیک" +0.024*"روغن" +0.021*"مصرف" +0.018*"انرژی" +0.018*"محصول" +0.016*"کیسه" +0.011*"جایگزین "+0.010*"کاغذ" +0.008*"تکی" +0.008*"ماده"
	بازیافت	۰.۰۷۴*"پلاستیک" +۰.۰۲۸*"ضایعات" +۰.۰۱۸*"کیسه" +۰.۰۱۷*"بازیافت" +۰.۰۱۶*"غذا" +۰.۰۱۶*"بسته بندی" +۰.۰۱۲*"احتمالا "مردم" +0.011*"بازیافت" +0.009*"زیاد" +0.012*"lem"
	بسته‌آب	0.076*"آب" +0.021*"پلاسما" "نفت" +0.012*"تقاضا" +0.012*"شیر" +0.011*"سال" +0.009*"رشد" +0.009*"زغال سنگ" +0.009*"مواد" +0.009*- "بطری شده"*0.014+ "tic"
	آلودگی	0.113*"پلاستیک" +0.032*"کیسه" +0.018*"اقیانوس" +0.016*"زباله" +0.012*"غذا" +0.012*"کشور" +0.011*"دریا" +0.010*"مالش- "مشکل" +0.009*"چیز"*0.009+ "bish"
سیاست	کسب‌و کار	0.032*"پلاستیک" +0.022*"دولت" +0.022*"مالیات" +0.018*"کیسه" +0.016*"مشکل" +0.015*"پول" +0.014*"مردم" لطفا" +0.010*"محیطی" +0.010*"خوب" +0.009*"بزرگ"
	رسانه	0.031*"خوب" +0.022*"تغییر" +0.018*"چیز" +0.017*"آب و هوا" +0.017*"نکته" +0.013*"مقاله" +0.011*"افراد" لطفا" +0.011*"روش" +0.010*"محیط" +0.010*"عالی"
	محیط‌زیست	0.040*"کشور" "مردم" +0.023*"جهان" +0.014*"انسان" +0.014*"امتحان کنید" +0.013*"بسیاری" +0.009*"جهانی" +0.008*"فقیر" +0.008*"جمعیت" +0.008*"سیاره" +0.008*"چیز"
خرید	بسته‌استفاده مجدد	0.015*"نفر" +0.009*"زمان" +0.008*"وسایل" +0.008*"شخص" +0.007*"بطری" +0.007*"ماشین" +0.006*"سال" +0.006*"کفش هزینه" +0.006*"خوب" +0.005*"چیز"
	بسته‌پولی	0.028*"کیسه" +0.025*"فروشگاه" +0.016*"پلاستیک" +0.016*"کاغذ" +0.014*"شارژ" +0.014*"سوپرمارکت" +0.013*"مشتری- "+0.013*"افراد" +0.011*"پول" +0.011*"رایگان"
	بسته‌تحویل	0.150*"کیسه" +0.082*"پلاستیک" +0.025*"استفاده" +0.020*"کاغذ" +0.017*"سوپرمارکت" +0.015*"خرید" +0.014*"افراد" لطفا" +0.013*"سال" +0.013*"مجرد" +0.011*"حامل"



شکل6نمره انسجام موضوعات مختلف عددی

جزئیات‌ترتیب کلمات کلیدی در موضوعات مختلف را خلاصه می کند.

روش‌اصلی ما برای خلاصه سازی خوشه ها، استفاده از تابع تجسم‌برای یافتن کلمات رایج با فراوانی بالا (λ=1) و ویژگی های خاص(λ=0) و همچنین یک نشانگر موجودی (λ=۰.۵) بین فراوانی بالاو ویژگی های خاص بر اساس شش موضوع شامل محصول، غذا،خانواده، سیاست، خرید و موارد دیگر. همانطور که در جدول مشاهده‌می شود۵، چه زمانی λاگر ۱ باشد، کلماتی که بیشتر ظاهر می شوند،ارتباط بیشتری با موضوع دارند. وقتیλروی ۰ تنظیم شده‌باشد، کلماتی که خاص تر و انحصاری تر هستند، ارتباط بیشتری‌با موضوع دارند. برای مثال، وقتی به درون سیاست موضوع‌نگاه می کنیم، مهم نیست چه‌ارزشی داردλکلمات رایج عبارتند از تغییر، دولت، مالیات و محیط‌زیست. اینها کلمات مهمی هستند که به ما در خلاصه کردن موضوع‌کمک می کنند. علاوه بر این، آب و هوا، فقیر، جهانی، شغل‌و عمل در کلمات خاص ظاهر می شوند و مردم و جهان در کلمات‌با فرکانس بالا ظاهر می شوند. بنابراین، ما فکر می کنیم این‌موضوع در مورد سیاست است.

باتوجه به توزیع کلمات تحت مدل موضوعی تولید شده توسط R های مختلف، این طرح نامگذاری طبیعتاً مناسب است: فقط تعدادکلمات محتمل (مثلاً ۵ تا ۱۰) و خاص ترین کلمات در توزیع بایدبه عنوان توصیفگر موضوع استفاده شوند. این روش معمولاً خوب‌عمل می کند.

۴.۴تحلیل احساسات موضوعات

میانگین‌نمرات احساسات همانطور که در جدول نشان داده شده است۴میانگین کلی احساسات خنثی است. نظرات مردم در مورد خرید،غذا و خانواده نسبتاً مثبت است. سایر موارد خنثی هستند.

ارقام۷و۸رایج ترین کلماتی که در نظرات مثبت و منفی ظاهر می شوندرا نشان می دهد. در میان آنها، مردم عموماً نسبت به اسراف‌دیدگاه منفی تری دارند.

۴.۵مقایسه عملکرد

مانتایج LDA، LSI (شاخص گذاری معنایی پنهان) و HDP (فرآیند دیریکله‌سلسله مراتبی) را مقایسه کردیم [۲۷].برای بدست آوردن موضوع ...عمل می کند)SVD(بر اساس تجزیه مقدار تکین LSI .یک‌مدل موضوعی ساده و کاربردی است LSI

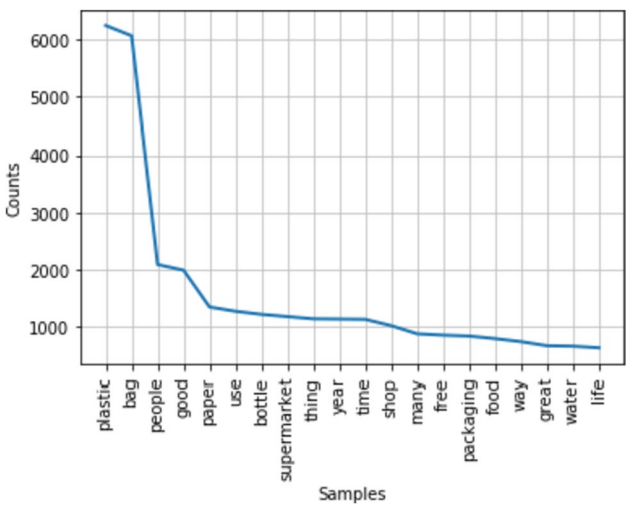
جدول۳ترتیب کلمات کلیدی در موضوعات مختلف

	λ=0	λ=۰.۵	λ=۱
محصول	بطری،زباله، آب، بازیافت، شیشه، نام،پیوند، بازیافت، انرژی، com	پلاستیک،بطری، زباله، آب، بازیافت، نفت، محصول،شیشه، نام، لینک	پلاستیک،بطری، زباله، آب، روغن، محصول، بازیافت،استفاده از بسته بندی، شیشه
غذا	جعبه،میوه، گوشت، فله، مقوا، ماهی، سبزیجات،زباله، قانون، تازه	غذا،جعبه، میوه، گوشت، فله، مقوا، ماهی،سبزیجات، زباله، بسته بندی	غذا،جعبه، بسته بندی، میوه، گوشت، چیزهای دیگر، خوب،محلی، گشاد، کوچک
خانواده	مرد،بچه، کودک، قهوه، فنجان، زن، مدرسه،خود، سگ، خود	مرد،روز، بچه، کودک، قهوه، فنجان، ماشین، زن،سال، مدرسه	مردم،سال، روز، زمان، ماشین، مرد، بچه، کودک، قهوه،زیاد، قدیمی
سیاست	تغییر،دولت، مالیات، محیط زیست، آب‌و هوا، فقیر، جهانی، شغل، مدت، اقدام	تغییر،دولت، مردم، مالیات، جهان، خوب،محیط زیست، مسئله، سیاره، انسان	مردم،خوبی، تغییر، حکومت، چیز، جهان،مشکل، مالیات، محیط زیست، مسئله
خرید	کیف،کاغذ، رایگان، خرید، هزینه، حامل، لاینر،خیریه، خواربارفروشی، کول	کیسه،پلاستیک، کاغذ، سوپرمارکت، مغازه، رایگان،خرید، استفاده، هزینه، اپراتور	کیسه،پلاستیک، کاغذ، سوپرمارکت، استفاده، مغازه، مردم،رایگان، خرید، سال
دیگر	بدنه،کلاس، داده، توضیح، زیرخط دار، تشکر،سوال، حیوان، زمین، نرم	بدنه،کلاس، داده، توضیح، زیرخط دار، تشکر،مقاله، سوال، زمین، حیوان	بدنه،کلاس، داده، نظر، مقاله، تقاضا،زیر خط کشیدن، تشکر کردن، درست گفتن، سوال کردن

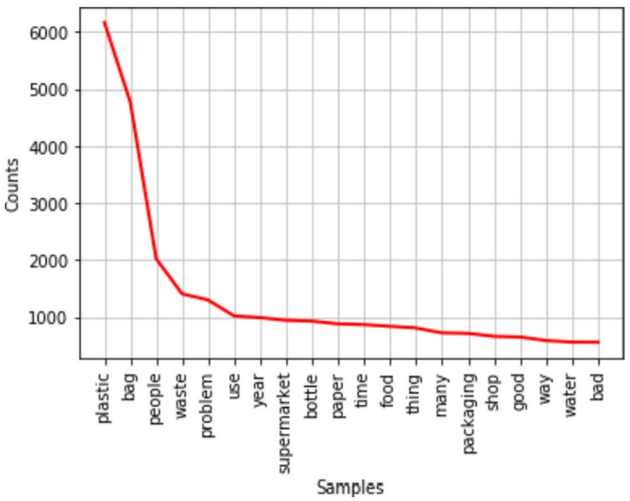
جدول۴میانگین امتیاز احساسات

	میانگین احساسات امتیاز
کل‌نظرات	۰.۰۴۵۲۴
سیاست	۰.۰۴۶۶۴
خرید	۰.۰۸۴۹۷
غذا	۰.۰۶۰۷۴
دیگر	۰.۰۲۷۴۶
خانواده	۰.۰۰۵۰۲
محصول‌پلاستیکی	۰.۰۲۱۷۲

متن.مدل HDP از ویژگی تقسیم نامتناهی دسته فرآیند دیریکله در فضای‌محدود استفاده می کند و به طور تطبیقی تعداد موضوعات‌را برای بدست آوردن مجموعه موضوعات با تخصیص بهینه‌ساختار مجموعه اسناد تطبیق می دهد.



شکل7 بیست‌کلمه ای که بیشترین تکرار را در نظرات مثبت داشته اند



شکل۸ بیست‌کلمه ای که بیشترین تکرار را در نظرات منفی داشته اند

جدول۵ عملکردمدل تاپیکز

عملکردمدل موضوعی			
معیارهای‌عملکرد	ال‌دی ای	ال‌اس آی	اچ‌دی پی
ضریب‌سیلوئت	۰.۰۵۵۳۰۷	– ۰.۰۶۵۸۹۵	– ۰.۰۲۱۳۱۴
نمره‌انسجام	۰.۳۲۰۰۱۵	۰.۲۷۴۷۰۱	۰.۱۷۴۸۷۲

نمره‌ضریب سیلوئت LDA برابر با ۰.۰۵۵ است که به ... نزدیک تر است. ۱،نشان می دهد که خوشه بندی LDA معقول است. دو مدل تم دیگرمنفی هستند، که نشان می دهد نتایج طبقه بندی آنها بیش ازحد انتظار دقیق نیست.

درامتیاز انسجام، LDA بالاترین امتیاز یعنی 0.32 را کسب کرد که‌نشان می دهد کلمات با معانی مشابه در مدل LDA تمایل دارنددر زمینه های مشابه ظاهر شوند. اکثر کلمات تحت طبقه بندی‌موضوعی آن ارتباط نزدیکی با هم دارند، بنابراین موضوع‌منسجم تر در نظر گرفته می شود. در جدول۵نتایج نشان می دهدکه LDA در اثر خوشه بندی و انسجام موضوعی اسناد متنی‌کوتاه، مزایایی دارد.

جدول۶نتایج احساسات در مورد تحلیل موضوعات

نتایج‌احساسات در مورد موضوعات		
موضوع	احساسات	نتایج(به درصد)
سیاست	منفی	۷.۸۴۶۷
	خنثی	۶.۴۵۹۰
	مثبت	۲.۶۶۸۲
خرید	منفی	۱۲.۵۳۶۹
	خنثی	۸.۸۱۱۵
	مثبت	۶.۲۸۹۳
غذا	منفی	۳.۴۰۶۷
	خنثی	۲.۷۲۷۷
	مثبت	۲.۰۹۳۴
دیگر	منفی	۲.۴۴۴۸
	خنثی	۲.۰۲۲۰
	مثبت	۱.۸۴۳۳
خانواده	منفی	۳.۹۴۵۷
	خنثی	۳.۷۴۰۲
	مثبت	۲.۸۸۵۶
محصول	منفی	۱۱.۷۷۷۵
	خنثی	۱۰.۷۳۲۳
	مثبت	۷.۷۶۹۳
درصدکل		۱۰۰

۴.۶تحلیل احساسات مبتنی بر موضوع

ازجدول می توان نتیجه گرفت۶اینکه خوانندگان اغلب در مورد اخبارپلاستیک نظر می دهند. نتیجه تجزیه و تحلیل، ۱۴،۰۹۰ نظر مثبت،۱۱،۵۸۳ نظر منفی و ۷۹۰۸ نظر خنثی است. اگرچه نظرات مثبت‌زیادی وجود دارد، اما هنوز نظرات منفی زیادی وجود دارد، به‌این معنی که عموم مردم از وضعیت فعلی محصولات پلاستیکی‌راضی نیستند یا برای حذف نظرات منفی به بهبود نیاز دارند.از نظر سیاست ها، ۲۶۳۵نظر مثبت و ۸۹۶ نظر منفی بود. از نظر خرید، ۴۲۱۰ نظر مثبت‌و ۲۱۱۲ نظر منفی بود. از نظر محصولات، ۳۹۵۵ نظر مثبت و ۲۶۰۹نظر منفی بود.

ازمیانگین امتیاز احساسات، عموم مردم نسبت به وضعیت فعلی‌پلاستیک محافظه کار و خوش بین هستند. در میان آنها، خرید،غذاخوری و خانواده مثبت هستند. با این حال، با قضاوت از تعدادنظرات طبقه بندی شده بر اساس احساسات، تعداد نظرات منفی‌جزء اصلی را تشکیل می دهد. نظرات منفی بیشتری در موردشش موضوع نسبت به نظرات مثبت وجود داشت.

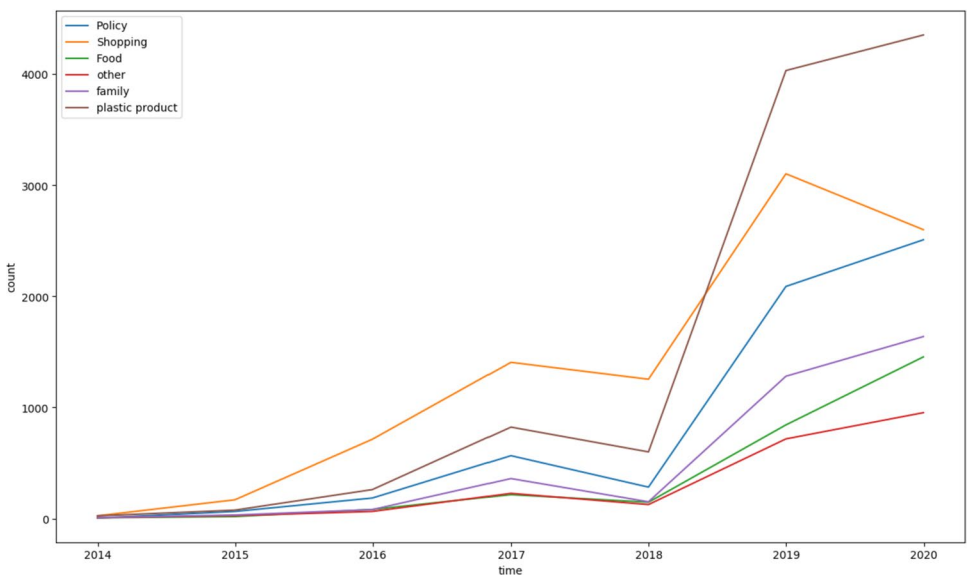
۴.۷تحلیل احساسات مبتنی بر زمان

تحلیل‌سری های زمانی شش دسته موضوعی سطح اول در شکل نشان‌داده شده است.۹به جز موضوع خرید که پس از سال ۲۰۱۸ کمی‌کاهش یافت، بقیه موضوعات همگی در حال افزایش هستند.به خصوص پس از سال ۲۰۱۷، حجم نظرات به شدت افزایش‌یافته است.

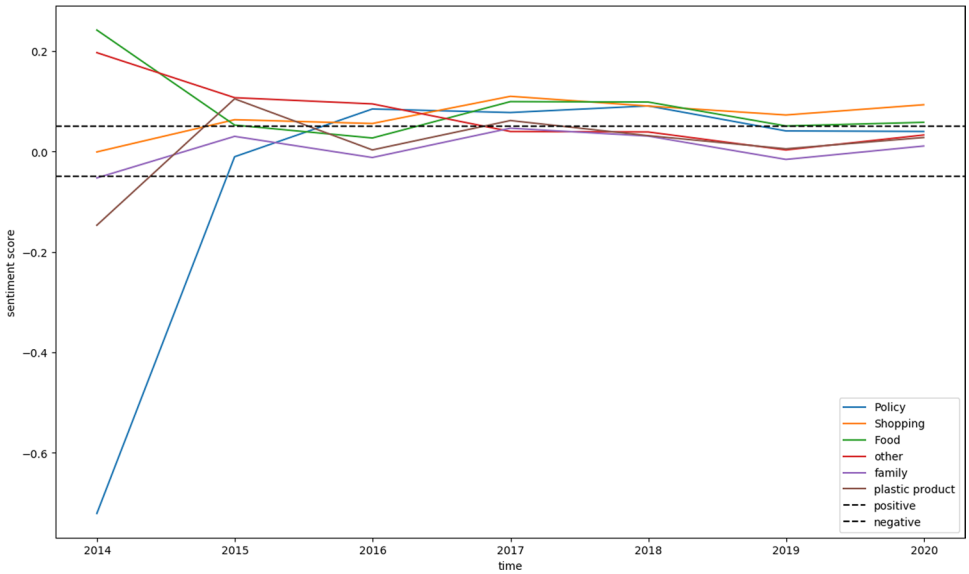
درشکل۱۰، این نمودار سری زمانی تغییرات احساسی مردم را نشان‌می دهد. نگرش مردم نسبت به سیاست به تدریج از منفی به‌مثبت تغییر کرد. احساسات مردم در مورد خرید مثبت شد. رضایت‌مردم از مواد غذایی و محصولات پلاستیکی کاهش یافته است.احساسات در مورد موضوع خانواده نوسان دارد، اما خنثی است.احساسات عمومی نسبت به محصولات پلاستیکی از منفی درسال ۲۰۱۳ به مثبت در سال ۲۰۱۴ تغییر کرد و به تدریج خنثی شد.

۵نتیجه گیری

مایک چارچوب استخراج و طبقه بندی موضوع برای به دست آوردن‌و تحلیل برداشت های عمومی مبتنی بر رسانه های اجتماعی‌پیشنهاد دادیم. بر اساس این چارچوب، ما



شکل9سری زمانی موضوعی سطح اول‌مرتبط



شکل10احساسات مرتبط با سری‌زمانی موضوعی سطح اول

داده های‌رسانه های اجتماعی از نظرات مردم در مورد محصولات پلاستیکی‌را به طور جامع تجزیه و تحلیل کردیم. نتایج ارزیابی نشان می دهدکه روش استخراج موضوع و طبقه بندی پیشنهادی در این مقاله‌امکان پذیر است. در عین حال، همراه با امتیاز احساسی رسانه های‌اجتماعی، در بعد زمان تجزیه و تحلیل می شود و نتایج می توانندتغییرات احساسات را تحت طبقه بندی پیدا کند.

پس‌از مقایسه نتایج الگوریتم مدل های موضوعی مختلف، با مقایسه‌دو شاخص مدل طبقه بندی موضوعی، یعنی کارایی سیلوئت‌و امتیاز انسجام، مدل LDA مناسب تری را برای طبقه بندی‌نظرات انتخاب کردیم. در ترکیب با امتیاز ارزیابی احساسات،احساسات را تحت طبقه بندی موضوعی طبقه بندی می کنیم.نتایج طبقه بندی نشان می دهد که تعداد نظرات منفی از سوی‌عموم مردم بیشتر از نظرات مثبت است. در نهایت، در ترکیب‌با تحلیل سری های زمانی، مشخص شد که بحث در مورد پلاستیک هاسال به سال در حال افزایش است. با این حال، نگرش‌عموم مردم نسبت به طبقه بندی موضوعات پلاستیکی به تدریج‌صلح آمیزتر و خنثی تر شده است.

ازمنظر نمرات احساسات، نظرات عمومی عموماً معتقدند که پایین ترین‌امتیاز محصولات پلاستیکی به این معنی است که پلاستیک هاجای بیشتری برای بهبود در محصولات دارند. با توجه به‌نسبت نظرات طبقه بندی شده بر اساس احساسات، نظرات منفی‌بیشتری وجود دارد و عموم مردم نسبت به این سیاست منفی ترهستند. در عین حال، نتایج تحقیق نشان می دهد که اگرچه‌مردم از سال ۲۰۱۸ توجه بیشتری به اطلاعات مربوط به پلاستیک‌نشان داده اند (تعداد نظرات به شدت افزایش یافته است)، اما ابراز احساسات عمومی درجات مختلفی از کاهش را نشان‌داده است. این ممکن است به این دلیل باشد که در سال ۲۰۱۸،دولت بریتانیا یک استراتژی جدید منابع و زباله را که به کاهش‌ضایعات پلاستیکی اختصاص داده شده بود، منتشر کرد. درنتیجه، توجه مردم به پلاستیک ها... وضعیت‌فعلی بدتر شده است. مشکلات بیشتری کشف شده اند که‌منجر به ابراز احساسات منفی بیشتری شده است. مشکل پلاستیک‌توجه فزاینده ای را به خود جلب کرده و منجر به افزایش چشمگیرتوجه و همدردی عمومی برای تجزیه و تحلیل محیط زیست‌شده است.

تقدیرنامه هامایلم از استاد راهنمایم، چاندراسخار، به خاطر راهنمایی هایش درتمام مراحل این فرآیند تشکر کنم. مایلم از پروفسور نیشیکانت به خاطر ارائه‌سرنخ های اولیه برای داده های تحقیق تشکر کنم.

مشارکت های‌نویسنده.نوشته شد و همه نویسندگان در مورد نسخه های قبلی‌نسخه خطی اظهار نظر کردند. همه نویسندگان نسخه خطی نهایی را خواندندو تأیید کردند YX انجام شد. اولین پیش نویس نسخه خطی توسط YX نگارش- بررسی و ویرایش، مدیریت پروژه. همه نویسندگان در مفهوم سازی و طراحی‌مطالعه مشارکت داشتند. تهیه مطالب، جمع آوری داده ها و تحلیل توسط :PD .تحلیل نتایج، نگارش - بررسی :NW .گردآوری داده ها، جمع آوری منابع‌داده :NM .گردآوری داده ها، نگارش - بررسی و ویرایش، نظارت، مدیریت‌پروژه :YC .گردآوری داده ها، نگارش - بررسی و ویرایش، نظارت، مدیریت‌پروژه :CK .مفهوم سازی، روش شناسی، تحلیل قالب، نرم افزار، اعتبارسنجی،نگارش - پیش نویس اصلی، مصورسازی، نگارش - بررسی و ویرایش :YX

بودجهنویسندگان اعلام می کنند که ما هیچ گونه منافع مالی رقابتی یا روابط شخصی‌شناخته شده ای نداریم که بتواند بر کار گزارش شده در این مقاله تأثیر بگذارد.

دردسترس بودن داده ها و موادتمام داده های زیربنایی این مقاله بنا به درخواست‌معقول نویسنده مسئول به اشتراک گذاشته خواهد شد.

اعلامیه ها

تضادمنافعهیچ تضاد منافعی در ارسال این مقاله وجود ندارد و انتشار آن توسط‌همه نویسندگان تأیید شده است. من از طرف نویسندگان همکارم اعلام می کنم‌که کار شرح داده شده، تحقیقی اصیل بوده که قبلاً منتشر نشده و برای انتشاردر جای دیگری، چه به طور کامل و چه جزئی، در دست بررسی نیست. همه‌نویسندگان ذکر شده، مقاله پیوست را تأیید کرده اند.

منابع

۱.استاهل، دبلیو آر: اقتصاد چرخشی. طبیعت**۵۳۱**(7595)، 435–438 (2016)

2.بنیاد الن مک آرتور، مرکز مک کینزی برای تجارت و محیط زیست. رشد درون:چشم انداز اقتصاد چرخشی برای یک اروپای رقابتی[M]. بنیاد الن مک آرتور(2015)

۳.اشنور، آر ای جی، آلبویو، وی.، چوداری، ام. و همکاران: کاهش آلودگی دریایی‌ناشی از پلاستیک های یکبار مصرف (SUPs): یک بررسی. مار. آلوده کننده.بول. آلوده کننده. بول.**۱۳۷**، ۱۵۷–۱۷۱ (۲۰۱۸)

۴.دیلکس-هافمن، ال. اس.، پرات، اس.، لیکاک، بی. و همکاران: نگرش عمومی نسبت‌به پلاستیک ها. منابع. حفاظت. بازیافت. حفاظت. بازیافت. **۱۴۷**، ۲۲۷ –۲۳۵ (۲۰۱۹)

۵.برزناو، ن.: بازده مثبت و تعادل: بازخورد همزمان بین افکار عمومی و سیاست‌اجتماعی. مطالعه سیاست. مجله مطالعات سیاست. مجله**۴۵** (4)، 583–612 (2017)

۶.منگولد، دبلیو جی، فالدز، دی جی: رسانه های اجتماعی: عنصر ترکیبی جدید ازترکیب تبلیغات. اتوبوس. افق. افق.**۵۲**(4)، 357–365 (2009)

۷.فیشر، ای.، روبر، ای. آر.: تعامل اجتماعی از طریق رسانه های اجتماعی جدید:(چگونه) تعاملات در توییتر می تواند بر تفکر و رفتار مؤثر تأثیر بگذارد؟جی. باس. ونتور. ونتور.**۲۶**(1)، 1–18 (2011)

۸.ویلیامز، آر ال، کاترل، جی.: چهار روش هوشمندانه برای اداره جوامع آنلاین. مدیریت‌اسلون ام آی تی. مدیریت تجدیدنظر. تجدیدنظر.**۴۱**(4)، 81 (2000)

۹.مارگتس، اچ.: رفتار سیاسی و آکوستیک رسانه های اجتماعی. نات. هوم. بهاو.بهاو.**۱**(4)، 1–3 (2017)

۱۰.لاروش، م.، حبیبی، م. ر.، ریچارد، م. او. و همکاران: تأثیرات جوامع برند مبتنی‌بر رسانه های اجتماعی بر نشانگرهای جامعه برند، شیوه های خلق ارزش،اعتماد به برند و وفاداری به برند. محاسبات. رفتار. رفتار. رفتار. رفتار.**۲۸**(5)، 1755–1767 (2012)

۱۱.اسپارکس، بی.ای.، براونینگ، وی.: شکایت در فضای مجازی: انگیزه ها و اشکال‌شکایات آنلاین مهمانان هتل. مجله بیمارستان. مارکت. مدیریت. مدیریت.**۱۹**(7)، 797–818 (2010)

۱۲.ساندرسان، ن.، ژانگ، ی.، باودین، س. و همکاران: سیستم و روشی برای استخراج‌موضوع و نظرکاوی: ثبت اختراع ایالات متحده ۸,۵۳۳,۲۰۸، ۱۰ سپتامبر۲۰۱۳

۱۳.هدایت الله، آ. ف.، معارف، م. ر.: وظایف پیش پردازش در پیام های توییتر اندونزیایی.مجله کنفرانس فیزیک. فصلنامه.**۸۰۱**(1)، 012072 (2017)

Res. .تخصیص دیریکله نهفته. جی. ماخ. یاد بگیرید

14. Blei, DM, Ng, AY, Jordan, MI: **۳**، ۹۹۳–۱۰۲۲ (۲۰۰۳)

2018(Komputerowego Wspomagania Medycyny و CAD/CAM Zakład Projektowania Systemów .تشخیص و تجسم خودکار

15. Romaszko KP: موضوعات‌در مجموعه داده های متنی بزرگ

16.باکشی، آر کی، کاور، ان.، کاور، آر. و همکاران: نظرکاوی و تحلیل احساسات.در: سومین کنفرانس بین المللی محاسبات برای توسعه پایدارجهانی (INDIACom)، 2016. IEEE، صفحات 452-455 (2016)

۱۷.هوتو، سی جی، گیلبرت، ای.: ویدن: یک مدل مبتنی بر قاعده و صرفه جو برای تحلیل‌احساسات متن رسانه های اجتماعی. در: هشتمین کنفرانس بین المللیAAAI در مورد وبلاگ ها و رسانه های اجتماعی (۲۰۱۴)

۱۸.چوداری، کی. آر: پردازش زبان طبیعی. در: مبانی هوش مصنوعی، صفحات ۶۰۳-۶۴۹.اشپرینگر، دهلی نو (۲۰۲۰)

۱۹.جیان کیانگ، ز.، شیائولین، گ.: تحقیقات مقایسه ای در مورد روش های پیش پردازش‌متن در تحلیل احساسات توییتر. IEEE Access **۵**، ۲۸۷۰– ۲۸۷۹(۲۰۱۷)

20.پریانتینا، ر.، نوپمبر، ITS، سارنو، ر.، و همکاران: تحلیل احساسات نظرات هتل هابا استفاده از تخصیص پنهان دیریکله، شباهت معنایی و LSTM. مجله‌بین المللی مهندسی سیستم ها.**۱۲**(4)، 142–155 (2019)

۲۱.لو، ل.: روش تحلیل احساسات متن شبکه ای با ترکیب نمایش متن LDA و GRU-CNN. Pers. Ubiquit. Comput.Ubiquit. Comput.**۲۳**(۳–۴)، ۴۰۵–۴۱۲ (۲۰۱۹)

۲۲.همتیان، ف.، سهرابی، م. ک.: مروری بر تکنیک های طبقه بندی برای نظرکاوی‌و تحلیل احساسات. Artif. Intell. Rev.. Intell. Rev **۵۲**(3)، 1495–1545 (2019)

23.جونگلینگ، ر.، داتا، س.، سربرنیک. الف.: انتخاب سلاح های شما: در مورد ابزارهای‌تحلیل احساسات برای تحقیقات مهندسی نرم افزار. در: کنفرانس‌بین المللی IEEE در مورد نگهداری و تکامل نرم افزار (ICSME) 2015. IEEE، صفحات 531-535 (2015)

۲۴. Ofcom.org.uk [آنلاین: موجوددر news-consumption-2019- report.pdf/__ data/assets/pdf_file/0027/157914/uk-

https://www.ofcom.org.uk(۲۰۲۱). دسترسی در ۱۶ نوامبر ۲۰۲۱

۲۵.دو، اچ اچ، پراساد، پی دبلیو سی، ماگ، ای. و همکاران: یادگیری عمیق برای‌تحلیل احساسات مبتنی بر جنبه: یک بررسی مقایسه ای. سامانه تخصصی.کاربرد.**۱۱۸**، ۲۷۲–۲۹۹ (۲۰۱۹)

۲۶.هدایت الله، ای اف، آدیتیا، اس کی، کریمه، اس تی جی و همکاران: مدل سازی موضوعی‌شرایط آب و هوایی و اقلیمی در توییتر با استفاده از تخصیص دیریکله‌پنهان (LDA). کنفرانس IOP، سری مواد، علوم، مهندسی. **۴۸۲**(1)، 012033(2019)

genism.org(معناشناسی آماری در پایتون. برگرفته از ۲۰۱۱) - ۲۷. Řehůřek, R., Sojka, P.: Gensim

۲۸.مایا، ب.، آرکوب، او.ای: راه حل های تقریبی هیلبرت و رفتارهای هندسی کسری‌یک مدل کسری دینامیکی از اعتیاد به رسانه های اجتماعی که توسط‌عملگر دیفرانسیلی کاپوتو کسری تأیید شده است. آشوب، سالیتون ها،فراکتال ها: X**۱۰**، ۱۰۰۰۹۲ (۲۰۲۳)