Archive of SID

عقیده‌کاوی در زبان فارسی مبتنی بر یادگیری انتقالی

سعید دهقانی اشکذری[۱]، دانشجوی دکتری؛ ولی درهمی[۲]، دانشیار؛ علی‌محمد زارع بیدکی[۳]، دانشیار؛ محمداحسان بصیری[۴]، استادیار

۱- دانشکده مهندسی کامپیوتر - دانشگاه یزد - یزد - ایران – sdehghani@stu.yazd.ac.ir
۲- دانشکده مهندسی کامپیوتر - دانشگاه یزد - یزد - ایران – vderhami@yazd.ac.ir
۳- دانشکده مهندسی کامپیوتر - دانشگاه یزد - یزد - ایران – alizareh@yazd.ac.ir
۴- دانشکده مهندسی کامپیوتر - دانشگاه شهرکرد – شهرکرد - ایران – basiri@sku.ac.ir

چکیده: در دهه گذشته بررسی عقاید، احساسات و هم‌چنین گرایش‌های انسان‌ها نقش موثری در تصمیم‌گیری مدیران و افراد داشته است. الگوریتم‌های یادگیری ماشین نقش مهمی در زمینه عقیده‌کاوی دارند، اما از یک مشکل بزرگ رنج می‌برند: اغلب الگوریتم‌های یادگیری ماشین فرض می‌کنند که ابعاد ویژگی و توزیع داده‌ها یکسان است، اما بسیاری از کاربردهای واقعی از این فرضیات تبعیت نمی‌کنند. در واقع، داده‌هایی که الگوریتم در آینده دریافت می‌کند ممکن است دارای ابعاد و ویژگی متفاوت و یا از توزیع دیگری باشند. در این مقاله با استفاده از یادگیری انتقالی و با تاکید بر انتقال ویژگی، روشی نوین را برای بهبود استخراج احساسات و مفاهیم موجود در عقاید ارائه می‌دهیم. در روش پیشنهادی، ابتدا ویژگی یا موضوع عقاید در دامنه زبانی مبدأ شناسایی شده و با جمع‌آوری صفات، قیدها و به‌طورکلی بسته‌ای از احتمالاتی که ممکن است برای یک ویژگی رخ دهد و ترجمه آن به زبان مقصد، یادگیری از زبان مبدأ به زبان مقصد انتقال می‌یابد. بررسی روش پیشنهادی روی داده‌های موجود در فروشگاه اینترنتی آمازون به‌عنوان دامنه مبدأ نشان می‌دهد با ایجاد الگوی انتقال ویژگی روی عقاید به زبان انگلیسی، می‌توان قطبیت موجود در ۷۷ در صد نظرات که به زبان فارسی ثبت شده (در فروشگاه دیجی‌کالا) را کشف نمود که نسبت به روش‌های SCL، SFA و TCA به‌ترتیب ۹، ۵ و ۵ در صد افزایش بازدهی را نشان می‌دهد.

واژه‌های کلیدی: عقیده‌کاوی، یادگیری انتقالی، انتقال ویژگی، قطبیت.

Persian Opinion Mining based on Transfer Learning

Saeed Dehghani Ashkezari[1], PhD Student; Vali Derhami[2], Associate Professor; Ali Mohammad Zare Bidoki[3], Associate Professor; Mohammad Ehsan Basiri[4], Assistant Professor

1- Faculty of Computer Engineering, Yazd University, Yazd, Iran, sdehghani@stu.yazd.ac.ir
2- Faculty of Computer Engineering, Yazd University, Yazd, Iran, vderhami@yazd.ac.ir
3- Faculty of Computer Engineering, Yazd University, Yazd, Iran, alizreh@yazd.ac.ir
4- Faculty of Computer Engineering, Shahrekord University, Shahrekord, Iran, basiri@sku.ac.ir

Abstract: In the past decade, the study of human opinions, feelings and tendencies has been very effective in the decision-making of managers and individuals. Machine learning algorithms play an important role in the field of opinion mining, but they suffer from a big problem: most of the machine learning algorithms assume that the feature dimensions and data distribution are equal, but most of real-world applications don't follow these assumptions. In fact, the data that the algorithm will receive in the future may have different dimensions or distributions. In this article, a new method for improving sentiment analysis of opinions is proposed by the aid of feature-based transfer learning. In the proposed method, initially, the feature or topic of the opinion in the source language domain is identified. Then, by collecting adjectives, adverbs and totally a package of probabilities about that feature and by translating it into the target language, learning from the source language is transferred into the target language. An analysis of the proposed method on the data available at the Amazon store as the source domain indicates that by creating a pattern of feature transferring in English, the polarity of 77% of the opinions in Persian (recorded at the Digikala store) can be extracted that outperforms the SCL, SFA and TCA models with 9, 5 and 5 percent respectively.

Keywords: Opinion mining, Transfer learning, Feature transfer, Polarity.

ارسال مقاله: ۱۳۹۷/۰۹/۰۷
تاریخ اصلاح مقاله: ۱۳۹۸/۰۳/۲۹، ۱۳۹۸/۰۴/۱۶و ۱۳۹۸/۰۶/۱۲
تاریخ پذیرش مقاله: ۱۳۹۸/۰۶/۲۵
نام نویسنده مسئول: ولی درهمی
نشانی نویسنده مسئول: ایران - یزد - صفاییه - دانشگاه یزد - دانشکده مهندسی کامپیوتر.

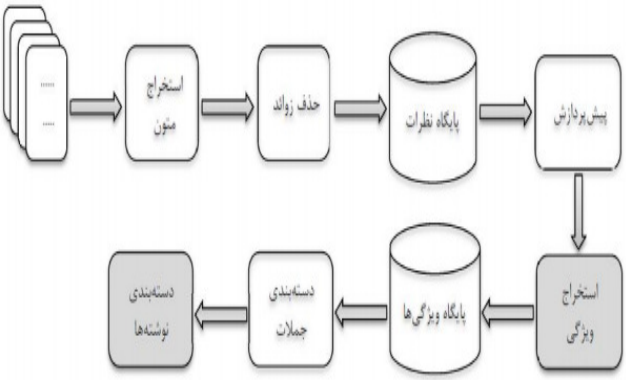
www.SID.ir

Archive of SID

۱- مقدمه

روزانه حجم عظیمی از عقاید توسط کاربران در وب و مخصوصاً در شبکه‌های اجتماعی منتشر می‌شود که به علّت حجم بسیار زیاد آن‌ها، تحلیل و پردازش انسانی عملا غیرممکن است. عقیده‌کاوی راه‌حلی برای تحلیل این عقاید به‌صورت خودکار است.

عقیده‌کاوی یا تحلیل احساسات زمینه مطالعاتی است که سعی می‌کند احساسات، رفتار، نظرات و تحلیل افراد مختلف را نسبت به یک موجودیت و ویژگی های آن بیان کند. این موجودیت می‌تواند یک محصول، خدمات، سازمان، فرد و سایر رخدادها و موضوعات باشد. به‌طورکلی عقیده‌کاوی شامل فرآیند کلی است که در شکل ۱ قابل‌مشاهده است و البته برخی از مراحل توسط بعضی از محققان نادیده گرفته می شود. مرحله پیش‌پردازش در این شکل شامل انتخاب اسناد مرتبط و مناسب، نشان‌گذاری، حذف ایست‌واژه‌ها، ریشه‌یابی و وزن‌دهی می‌شود که تاثیر مهم و موثری بر نتایج حاصل دارد [۱].



شکل ۱: مدل کلی عقیده‌کاوی

الگوریتم‌های یادگیری ماشین کمک شایانی به پیشرفت عقیده‌کاوی نموده‌اند، اما پیش‌فرض‌ها (یکسان بودن ابعاد ویژگی و توزیع داده‌ها) که نقش مهمی در الگوریتم‌های داده‌کاوی و یادگیری ماشین دارند، انعطاف‌پذیری را از این الگوریتم‌ها در حل بعضی از مسائل دنیای واقعی گرفته‌اند. معمولا در دنیای واقعی این فرضیات در نظر گرفته نمی شود. زمانی که نحوه توزیع داده‌ها متفاوت با شد، باید تمامی مدل‌های آماری با کمک داده‌های آموزشی جدید از ابتدا ساخته شوند که این کار بسیار هزینه‌بر خواهد بود. در چنین حالت‌هایی است که نیاز به استفاده از انتقال دانش به کمک یادگیری انتقالی به وجود می‌آید. یادگیری انتقالی، توانایی یک سیستم برای استخراج و اعمال دانش از کارهای قبلی به کار جدید است [۲]. در واقع یادگیری انتقالی یک مسئله تحقیقاتی در حوزه یادگیری ماشین است که تمرکز آن بر روی ذخیره کردن دانش به‌دست‌آمده از حل یک مسئله و به کار بردن آن برای حل مسئله متفاوت اما مرتبط دیگری است. برای مثال، دانش کسب شده ضمن یادگیری تشخیص خودروها (از روی تصاویر یا ویدئوها) قابل‌اعمال بر مساله تشخیص تراکتور نیز هست.

مطابق با [۳] یادگیری انتقالی به صورت زیر تعریف می‌شود:

اگر دامنه منبع D_S، کار منبع T_S، دامنه هدف D_T و کار هدف T_T باشد، هدف یادگیری انتقالی، بهبود یادگیری تابع پیش‌بینی هدف در D_T با استفاده از D_S و T_S است به‌طوری‌که $D_S \# D_T$ یا $T_S \# T_T$ باشند.

یکی از کاربردهای یادگیری انتقالی دسته‌بندی احساسات یا عقیده کاوی است. وظیفه یادگیری انتقالی در عقیده‌کاوی، دسته‌بندی خودکار عقاید در مورد یک محصول به سه دسته مثبت، منفی و خنثی از طریق آموزش دسته‌بندی عقاید در دامنه‌ای متفاوت است. برای انجام این کار، ابتدا باید مخزنی از عقاید برچسب‌گذاری شده را به‌عنوان دامنه مبدأ جمع‌آوری کرده و بعد از استخراج ویژگی‌ها و آموزش، برچسب عقاید را در دامنه مقصد مشخص کنیم. در یادگیری انتقالی مبتنی بر انتقال ویژگی، محققین فرض می‌کنند که یک فضای ویژگی مشترک که به‌وسیله نمونه‌های دامنه مبدأ و مقصد، به اشتراک گذاشته شده است وجود دارد و این فضای ویژگی می‌تواند به‌عنوان پلی برای انتقال دانش از دامنه مبدأ به دامنه مقصد به کار گرفته شود.

اغلب پژوهش‌های انجام‌شده در حوزه تحلیل احساسات و همچنین مجموعه‌داده‌های گردآوری‌شده بر روی زبان‌های انگلیسی، چینی و روسی صورت گرفته است و برخلاف این‌که زبان فارسی در حوزه وسیعی از آسیا صحبت و نوشته می شود، تلاش‌های پژوهشی نسبتا اندکی در زمینه عقیده‌کاوی و همچنین ایجاد و گردآوری مدل‌ها و جموعه‌داده‌های استاندارد در این زمینه صورت گرفته است [۴].

بر این اساس، هدف اصلی این مقاله، توسعه و بهبود الگوریتم‌های یادگیری انتقالی مبتنی بر ویژگی به‌منظور پردازش چنددامنه‌ای، رفع مشکل ویژگی‌های چندقطبی و اعمال دقیق‌تر عقیده‌کاوی در سطح جمله یا ویژگی است.

۲- کارهای مرتبط

اغلب تحقیقات درزمینه تحلیل احساسات در مورد عقاید انگلیسی صورت گرفته است، بنابراین اکثر ابزارها و داده‌های موجود به زبان انگلیسی هستند و بیشتر زبان‌های دنیا از فقدان داده‌های آموزشی مناسب رنج می‌برند. تحلیل احساسات بین‌زبانی مبحثی است در ارتباط با تطبیق دامنه که دو محرک اصلی برای انجام آن وجود دارد [۵، ۶]: نخست محققان در کشورهای مختلف می‌خواهند سیستم تحلیل احساسات را در زبان خود پیاده‌سازی کنند، درحالی‌که اکثر این تحقیقات به زبان انگلیسی انجام شده است و ممکن است ابزارهای مناسبی برای زبان‌های غیر انگلیسی وجود نداشته باشد. سؤال طبیعی در چنین مواردی این است که چگونه می‌توان از ابزارهای ترجمه و عقیده‌کاوی در زبان انگلیسی برای ایجاد سیستم مناسب عقیده‌کاوی در زبان‌های دیگر استفاده نمود. محرک دوم این است که بسیاری از شرکت‌ها می‌خواهند عقیده مشتریان خود از سایر کشورها را درباره محصولات و خدمات خود بدانند. اگر آن‌ها سیستم عقیده‌کاوی به انگلیسی داشته باشند می‌توانند به سادگی چنین سیستم‌هایی را در زبان‌های دیگر ایجاد نمایند. در زبان فارسی نیز تحقیقات کمتری برای عقیده‌کاوی صورت گرفته است.

www.SID.ir

Archive of SID

بسیاری از محققان این مسئله را بررسی کرده‌اند که اکثر آن‌ها در زمینه دسته‌بندی عقاید در سطح سند و جمله است. کارهای بسیار محدودی نظیر گوئو و همکاران [۷] در سطح ویژگی انجام شده است. ابتدا به برخی از پژوهش‌های انجام‌شده در سطح سند می‌پردازیم.

وان و همکاران [۸] از منابع عقیده‌کاوی در زبان انگلیسی برای دسته‌بندی احساسات در زبان چینی استفاده کردند. در نخستین گام، عقاید از زبان چینی به کمک چند مترجم به زبان انگلیسی ترجمه می‌شود. سپس از یک رویکرد مبتنی بر واژه‌نامه برای دسته‌بندی هر نسخه زبان انگلیسی استفاده می‌شود. این واژه‌نامه از یک مجموعه کلمات مثبت، منفی، منفی‌ساز و تشدیدکننده تشکیل شده است. سپس این الگوریتم امتیاز کلمات احساسی در هر عقیده را با توجه به منفی‌سازها و تشدیدکننده‌ها جمع‌بندی می‌کند. اگر امتیاز حاصل کوچک‌تر از صفر بود آن عقیده منفی و در غیر این صورت، مثبت در نظر گرفته می‌شود. برای دسته‌بندی نهایی هر عقیده، امتیاز نسخه‌های مختلف ترجمه با استفاده از روش‌های مختلف تجمیع مانند میانگین، ماکزیمم، میانگین وزن‌دار، رأی‌گیری و ... ترکیب می‌شود. اگر یک واژه‌نامه چینی نیز در دسترس باشد، تکنیک مشابهی را می‌توان برای نسخه چینی اعمال کرد و نتایج آن را با نتایج حاصل از نسخه‌های انگلیسی ترکیب نمود. نتایج نشان می‌دهد که تکنیک تجمیع تأثیر زیادی دارد. همچنین بروکه و همکاران [۹] نیز آزمایش مشابهی با ترجمه از زبان مبدأ (انگلیسی) به زبان مقصد (اسپانیولی) به کمک تنها یک مترجم انجام دادند و سپس از رویکرد مبتنی بر واژه‌نامه یا یادگیری ماشین برای دسته‌بندی اسناد زبان مقصد استفاده نمودند.

وان [۱۰] در پژوهش دیگری یک روش آموزش توأم پیشنهاد داد که از پیکره برچسب‌گذاری شده انگلیسی برای دسته‌بندی عقاید چینی به روش نظارتی استفاده می‌کند. در این پژوهش، هیچ منبعی از زبان چینی مورد استفاده قرار نمی‌گیرد. در آموزش، ورودی از یک مجموعه عقاید انگلیسی برچسب‌گذاری شده و عقاید چینی بدون برچسب تشکیل می‌شود. عقاید انگلیسی برچسب‌گذاری شده به زبان چینی و عقاید چینی بدون برچسب نیز به زبان انگلیسی ترجمه می‌شوند. بنابراین هر عقیده با یک نسخه انگلیسی و چینی همراه است. ویژگی‌های انگلیسی و چینی برای هر عقیده، به‌عنوان دو نمای مستقل و اضافی از عقاید در نظر گرفته می‌شود. سپس یک الگوریتم آموزش توأم با استفاده از SVM برای آموزش دو دسته‌بند مورد استفاده قرار می‌گیرد و در نهایت، این دو دسته‌بند با یکدیگر ترکیب می‌شوند. در مرحله دسته‌بندی، هر عقیده بدون برچسب چینی برای تست، ابتدا به زبان انگلیسی ترجمه شده و سپس دسته‌بند آموزش‌دیده برای دسته‌بندی عقاید به‌صورت مثبت و منفی اعمال می‌گردد.

وای و پال استفاده از روش یادگیری انتقالی را برای دسته‌بندی احساسات بین زبانی پیشنهاد کردند. با توجه به این واقعیت که ترجمه ماشین هنوز کامل نیست، برای کاهش نویز حاصل از ترجمه، آن‌ها به‌کارگیری روش SCL را برای یافتن مجموعه کوچکی از ویژگی‌های اصلی مشترک در هر دو زبان انگلیسی و چینی پیشنهاد دادند. برای کاهش مسئله پراکندگی داده‌ها و ویژگی‌ها نیز یک پرس‌وجو برای یافتن ویژگی‌های دارای همبستگی بیشتر با ویژگی‌های اصلی به موتور جستجو ارسال می‌شود و سپس ویژگی‌های کشف شده جدید برای ایجاد نمونه‌های بزرگ‌تر به‌کاربرده می‌شوند [۱۱].

گوئو و همکاران یک روش مبتنی بر موضوع را برای گروه‌بندی یک مجموعه از ویژگی‌های موجود در زبان‌های مختلف برای مقایسه مبتنی بر ویژگی عقاید از کشورهای مختلف پیشنهاد کرده‌اند [۷].

در تحقیقی از دوه و همکاران، این محققین ایده‌های خود را درباره تحقیق دسته‌بندی بین زبانی عقاید ارائه کردند. بر اساس این تجزیه‌وتحلیل، آن‌ها ادعا می‌کنند که عدم تناسب دامنه به‌وسیله خطاهای ترجمه ماشینی ایجاد نمی‌شود و کاهش دقت حتی در صورت ترجمه ماشینی کامل نیز اتفاق می‌افتد. آن‌ها همچنین بحث می‌کنند که مسئله تطبیق بین زبانی به‌طور کیفی متفاوت از مسائل تطبیق دامنه تک‌زبانی در پردازش زبان‌های طبیعی (NLP) است و بنابراین باید الگوریتم‌های جدید تطبیق دامنه مورد ملاحظه قرار گیرد [۱۲].

همانند دسته‌بندی عقاید چندزبانی در سطح سند، بعضی از محققین در سطح جمله نیز به این موضوع پرداخته‌اند. در اینجا نیز هدف استفاده از منابع و ابزارها در زبان انگلیسی و ترجمه‌های خودکار برای کمک در ایجاد سیستم عقیده‌کاوی در زبان‌های دیگر که منابع و ابزارهای کمتری دارند است. تحقیقات حاضر سه استراتژی اصلی را پیشنهاد می‌کنند:

۱. داده‌های تست موجود در زبان مقصد را به زبان مبدأ ترجمه و آن‌ها را به‌کمک دسته‌بند زبان مبدأ دسته‌بندی کنیم.
۲. پیکره آموزشی زبان مبدأ را به زبان مقصد ترجمه و یک دسته‌بند مبتنی بر پیکره در زبان مقصد ایجاد نماییم.
۳. یک عقیده یا احساس را از زبان مبدأ به زبان مقصد ترجمه و یک دسته‌بند مبتنی بر واژه‌نامه در زبان مقصد به کار ببریم.

در تحقیق دیگری از بانیا و همکاران، آزمایش‌های جامعی برای دسته‌بندی موضوعی بین‌زبانی در سطح جمله به‌وسیله ترجمه پیکره انگلیسی برچسب‌گذاری شده به پنج زبان دیگر انجام شده است. ابتدا نشان داده شد که استفاده از پیکره ترجمه‌شده برای آموزش تا حد قابل قبولی برای پنج زبان عمل می‌کند. ترکیب نسخه‌های ترجمه‌شده به زبان‌های مختلف با نسخه زبان انگلیسی اصلی برای تشکیل یک پیکره آموزشی واحد می‌تواند دسته‌بندی موضوعی پایه انگلیسی را بهبود دهد. دوم اینکه این پژوهش نشان داد که ترکیب پیش‌بینی‌های انجام‌شده توسط دسته‌بندهای تک‌زبانی با استفاده از رأی‌گیری اکثریت، قادر به تولید یک دسته‌بند موضوعی سطح جمله با دقت بالا است [۱۳].

تکنیک بوتین و همکاران نیز اسناد زبان مقصد را به انگلیسی ترجمه می‌کند و از یک روش مبتنی بر واژه‌نامه برای تعیین جهت‌دهی

www.SID.ir

Archive of SID

عقاید برای هر جمله حاوی یک موجودیت اســـتفاده می‌کند. این مقاله عملا در سطح ویژگی عمل می‌کند [۱۴].

در تحقیق لیو و همکاران یک مسـئله نسـبتاً متفاوت مورد بررسـی قرار گرفت. در این پژوهش فرض می شود که تعداد م شخ صی از عقاید برچ سب‌گذاری شده برای زبان‌های مبدأ و مق صد در د سترس ا ست و همچنین یک پیکره بدون برچ سب موازی وجود دارد. روش پی شنهادی می‌تواند به‌طور هم‌زمان دســـته‌بندی عقاید را برای هر دو زبان بهبود دهد و در واقع، یک الگوریتم EM مبتنی بر ماکزیمم آنتروپی اسـت که به‌طور هم‌زمان دو دسته‌بند تک‌زبانی را به‌طور هم‌زمان آموزش می‌دهد. در زمان آموزش، از این شهود ا ستفاده می شود که دو جمله یا سندی که موازی (به مفهوم ترجمه یکدیگر) هســـتند باید عقیده یکســـانی را نشان دهند [۱۵].

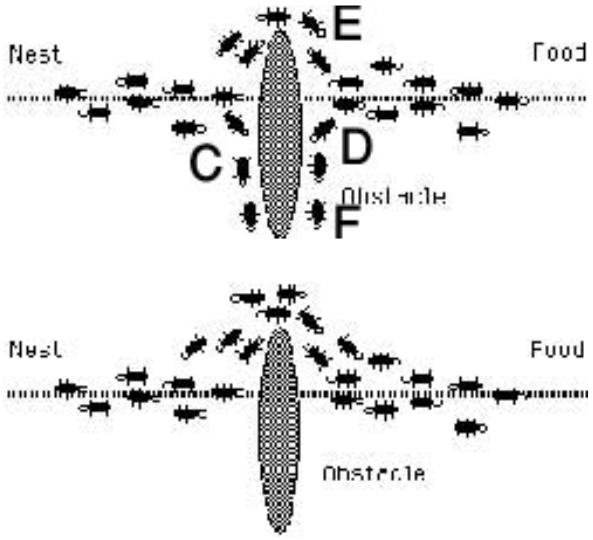
رســـولی برای انتقال زبانی بین دو زبان وقتی‌که اطلاعات چندانی از زبان مبدأ در دســـترس نیســـت، راهکاری مبتنی بر شـــبکه عصـــبی و اســـتفاده از اصـــطلاحات کلامی ارائه می‌دهد. رســـولی ســـیســـتم تجزیه‌وتحلیل احســـاســـات متقابل زبانی، بدون ترجمه ماشـــین و بدون اسـتفاده از داده‌های موازی باکیفیت مبتنی بر بررسـی داده‌های مرتبط در دامنه زبان مقصد و داده‌های موازی خارج از دامنه زبان مبدأ طراحی نمود. روش ابداعی رســـولی مشـــکل یادگیری‌های انتقالی که زبان مبدأ شان زبانی غیر از انگلیسی بود و فاقد دامنه دانش کمتری در مبدأ بودند را رفع نمود. رسـولی با بررسـی ۱۶ زبان هم‌خانواده نتیجه گرفت که روش پیشنهادیش می‌تواند سیستمی یادگیری انتقالی ایجاد کند که با دقت زیاد به کارایی یک سیستم تحت نظارت نزدیک گردد [۱۶].

در سال‌های اخیر توجه به عقیده‌کاوی در زبان‌های غیرانگلیسی نیز ر شد و تو سعه زیادی یافته ا ست که از میان آن‌ها می‌توان به زبان‌های فارسی (طبق جدول ۱)، عربی [۱۷]، چینی [۱۸]، چک [۱۹]، فرانسوی [۲۰]، آلمانی [۲۱]، هندی [۲۲]، ایتالیایی [۲۳]، ژاپنی [۲۴]، روســـی [۲۵]، اسپانیایی [۲۶] و تایلندی [۲۷] را نام برد.

۳- الگوریتم کلونی مورچگان

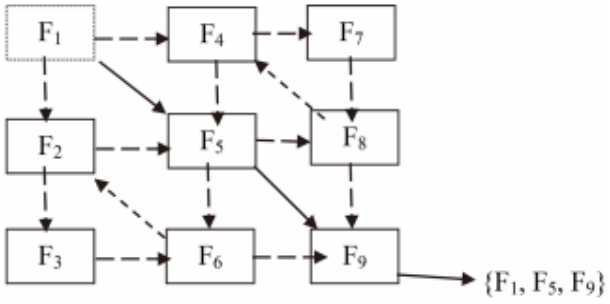
الگوریتم کلونی مورچه برای اولین بار توسط دوریگو[۱] و همکارانش در سال ۱۹۹۲به عنوان یک راه‌حل چندعامله[۲] برای مسائل سخت بهینه سازی مثل فروشنده دوره‌گرد[۳] ارائه شد [۲۸]. به گفته پژوهش‌گران مورچه‌ها نوعی هوش جمعی دارند. مشاهده ساختارهای اجتماعی آن‌ها نتیجه مبادلات چندجانبه‌ای است که به ظاهر ساده‌اند. این مبادلات به‌کمک ارتباط‌های شیمیایی صورت می‌گیرد که مورچه‌ها با شاخک های خود این علائم را دریافت می‌کنند. فرومون[۴] ماده‌ای شیمیایی است که مورچه‌ها پس از عبور از خود برجا می‌گذارند. آزمایشی ساده این پدیده را ثابت می‌کند. پژوهش‌گران با قرار دادن دو مسیر کوتاه و بلند در برابر مورچه‌ها برای رسیدن به غذا پی بردند پس از مدتی همه مورچه-ها از مسیرکوتاه‌تر عبور می‌کنند که علت آن تمرکز فرومون در مسیر آن‌ها است. هر چه میزان تمرکز فرومون درمسیر کوتاه‌تر بیشتر باشد میزان تردد در آن بیشتر می‌شود (شکل ۲).

در این روش، مورچه‌های مصنوعی به‌وسیله حرکت بر روی نمودار مسئله و با باقی گذاشتن نشانه‌هایی بر روی نمودار، همچون مورچه‌های واقعی که در مسیر حرکت خود نشانه‌هایی باقی می‌گذارند، باعث می شوند که مورچه‌های مصنوعی بعدی بتوانند راه‌حل‌های بهتری را برای مسئله فراهم نمایند [۲۹].



شکل ۲: انتخاب مسیر کوتاه‌تر توسط تمام مورچگان

انتخاب ویژگی یکی از کاربردهای مسائل زیرمجموعه‌ها[۵] است که می‌توان آن را همانند شکل ۳ به مساله‌ای مناسب برای حل توسط الگوریتم مورچگان تبدیل نمود:



شکل ۳: نمایش ACO مساله انتخاب ویژگی [۳۰]

$$P_i^k(t)=\begin{cases}\dfrac{[\tau_i(t)]^{\alpha}[\eta_i]^{\beta}}{\sum_{u\in J^k}[\tau_u(t)]^{\alpha}[\eta_u]^{\beta}} & i\in J^k \\ 0 & otherwise\end{cases} \qquad (۱)$$

که در آن، J^k مجموعه ویژگی‌های امکان‌پذیر است که می‌تواند به پاسخ جزیی اضافه شود؛ τ_i و η_i به‌ترتیب مقدار فرومون و میزان رضایت شهودی از ویژگی i را نشان می‌دهند. α و β نیز دو پارامتری هستند که اهمیت نسبی مقدار فرومون و اطلاعات شهودی را نشان می‌دهند و در واقع، تابع احتمال انتقال فوق تعادلی میان شدت فرومون τ_i (سابقه جرکت‌های موفق گذشته) و اطلاعات شهودی η_i (مطلوبیت حرکت) ایجاد می‌کند.

www.SID.ir

Archive of SID

جدول ۱: نمونه‌هایی از بررسی عقاید فارسی

مرجع	روش کار	مزایا	معایب
[۳۱]	ابتدا به تولید منبع واژگان فارسی موسوم به PersianClues و استفاده از آن در تعیین موضوع عقاید پرداخته می‌شود و سپس قطبیت هر سند به کمک الگوریتم یادگیری ماشین SVM به دست می آید.	ایجاد منبع واژگان جدید در زبان فارسی، ارائه روش نظارت‌نشده در عقیده‌کاوی فارسی، توسعه هم‌زمان موضوع و قطبیت	در نظر گرفتن اسناد به‌صورت بسته‌کلمات[۶]، استفاده از کلمات به‌عنوان ویژگی مورداستفاده در تعیین قطبیت
[۳۲]	چالش‌های عقیده‌کاوی در زبان فارسی شامل تنوع پسوندها، شکل‌های مختلف فاصله‌گذاری بین کلمات و ازدیاد کلمات محاوره‌ای مورد بررسی قرار می‌گیرد و سپس روش جدیدی برای استخراج ویژگی‌ها در جهت تعیین قطبیت اسناد فارسی ارائه می‌گردد	ارائه رویکرد جدیدی به نام MMI جهت استخراج ویژگی	عدم استفاده از لغت‌نامه، محدودیت تعداد نمونه‌ها برای آموزش، استفاده از یک نوع ویژگی یعنی n-gram، نیاز به انجام محاسبات بیشتر نسبت به روش‌های مبتنی بر لغت‌نامه
[۳۳]	ابتدا عملیات پیش‌پردازش شامل نرمال‌سازی کلمات، تصحیح غلط‌های املائی، ریشه‌یابی کلمات و تجزیه متن به جملات سازنده آن انجام می‌شود و سپس قطبیت جملات به‌کمک SentiStrength به دست می‌آید. در نهایت نیز امتیاز هر عقیده با استفاده از روش تجمیع دمپستر-شیفر[۷] به دست می‌آید.	در نظر گرفتن پیش‌پردازش‌های مختلف از قبیل تصحیح غلط‌های املایی و ریشه‌یابی کلمات، سفارشی‌سازی SentiStrength به‌منظور استفاده در زبان فارسی، ارائه رویکرد جدیدی مبتنی بر لغت‌نامه در عقیده‌کاوی زبان فارسی	عدم جامعیت روش پیشنهادی در ملاحظه منفی‌سازها، صفت‌ها و قیدها
[۳۴]	ابتدا SentiWordNet فارسی به‌کمک WordNet فارسی ایجاد می‌شود که از آن برای وزن‌دهی ویژگی‌ها استفاده می‌شود. سپس تاثیر این وزن‌دهی روی طبقه‌بندی اسناد با استفاده از روش‌های Logistic Regression، NB و SVM مورد بررسی قرار می‌گیرد	بررسی تاثیر وزن‌دهی ویژگی‌ها، بررسی تأثیر تعداد نمونه‌ها و مجموعه‌داده‌های نامتعادل در عقیده‌کاوی فارسی	عدم ارائه راه‌کاری جهت تصحیح غلط‌های املائی و تقطیع اسناد به جمله در عملیات پیش‌پردازش
[۳۵]	ابتدا به تهیه و جمع‌آوری لغت‌نامه فارسی از دو منبع برخط و سپس به ایجاد یک چهارچوب مبتنی بر GATE جهت پیش‌پردازش متون فارسی پرداخته می‌شود	ایجاد لغت‌نامه مخصوص زبان فارسی	محدودیت مجموعه‌داده‌های جمع‌آوری‌شده، لزوم در نظر گرفتن قواعد بیشتر و نوشتار اختصاری در زبان فارسی
[۳۶]	نخست لغت‌نامه عقیده‌کاوی برای زبان فارسی ایجاد می‌شود و بعد از انجام پیش‌پردازش، به استخراج ویژگی‌ها (با دو روش مبتنی بر تکرار و مبتنی بر دستور زبان) پرداخته می‌شود. در نهایت نیز مرحله پس‌پردازش جهت اصلاح ویژگی‌های حاصل از مرحله قبلی و تعیین قطبیت نهایی آن‌ها انجام می‌شود.	ایجاد لغت‌نامه عقیده‌کاوی برای زبان فارسی، عدم نیاز به داده‌های آموزشی، عدم وابستگی به دامنه خاص	عدم لحاظ کردن ارجاع مشترک[۸]، عدم استخراج ویژگی‌های ضمنی و غیرمستقیم، عدم تبدیل کامات محاوره‌ای به نوشتار رسمی
[۳۷]	نخست به بررسی تاثیر ابزارهای مختلف پردازش زبان‌های طبیعی (NLP)[۹] جهت پیش‌پردازش و رویکردهای مختلف دسته‌بندی احساسات پرداخته می‌شود و سپس یک لغت‌نامه WordNet جامع به نام FerdowsNet توسعه داده می‌شود.	ایجاد لغت‌نامه احساسی FerdowsNet و روش جدید PSWN[۱۰] به‌منظور استفاده از English SentiWordNet در زبان فارسی	محدودیت لغت‌نامه و پیکره[۱۱] مورداستفاده، انجام عقیده‌کاوی در سطح سند، تشخیص صرفا قطبیت عقاید

www.SID.ir

Archive of SID

بعد از این‌که تمام مورچه‌ها پاسخ خود را تکمیل کرده‌اند، تبخیر فرومون روی تمام گره‌ها طبق رابطه زیر اعمال می‌شود و هر کدام از مورچه‌ها مقداری از فرومون را در گره‌هایی که عبور کرده است به جا می‌گذارد:

$$\Delta\tau_i^k(t)=\begin{cases}\phi.\gamma(S^k(t))+\dfrac{\varphi.(n-|S^k(t)|)}{n} & i\in J^k\\ 0 & otherwise\end{cases} \quad (۲)$$

که در آن $S^k(t)$، زیرمجموعه‌ای از ویژگی‌ها است که توسط مورچه k در تکرار t یافت شده است و $|S^k(t)|$ نیز اندازه آن است. ϕ و φ نیز دو پارامتری هستند در محدود صفر تا یک که وزن نسبی عملکرد دسته‌بند و اندازه زیرمجوعه ویژگی‌ها را کنترل می‌کنند و $\varphi=1-\phi$. فرض ما بر این است که عمل‌کرد دسته‌بند دارای اهمیت بیشتری نسبت به طول زیرمجموعه است یعنی $\phi=0.8$ و $\varphi=0.2$ در نظر گرفته می‌شود.

در عمل، افزایش فرومون توسط مورچه‌ها و تبخیر آن به‌وسیله قاعده زیر روی تمام گره‌ها اعمال می‌شود:

$$\tau_i(t+1)=(1-\rho)\tau_i(t)+\sum_{k=1}^{m}\Delta\tau_i^k(t)+\Delta\tau_i^g(t) \quad (۳)$$

در این رابطه، m تعداد مورچه‌ها در هر تکرار، $\rho\in(0,1)$ ضریب کاهش فرومون و g نماینده بهترین مورچه در هر تکرار است [۳۸].

الگوریتم بهینه‌سازی مورچه‌ها متعلق به کلاس فراکتشافی است که شامل الگوریتم‌های تقریبی برای به دست آوردن راه‌حل کافی خوب در میزان زمان محاسباتی معقول، برای مسائل بهینه‌سازی ترکیبی سخت می‌باشد.

۴- روش پیشنهادی

در روش پیشنهادی ابتدا از طریق جدول، ایستواژه‌ها و حروف اضافه موجود در نظرات زبان مبدأ حذف می‌گردند. سپس با استفاده از جدول قواعد زبان مبدأ، جملات شناسایی گردیده و یک نظر به جملات سازنده آن تقسیم می‌گردد. با استفاده از لغت‌نامه سنتی‌وردنت [۳۹] نقش هر کلمه با توجه به مکان قرارگیری آن در جمله و نسبت به کلمات قبل و بعد آن مشخص می‌گردد. سپس از طریق الگوریتم کلونی مورچگان به تشخیص ویژگی در جملات یک نظر خواهیم پرداخت.

دسته نامزد[۱۲] ویژگی (در جملات یک نظر از زبان مبدأ) از بین کلماتی انتخاب خواهند شد که نقش‌شان در جمله اسم باشد. منظور از ویژگی در اینجا، ماهیت یا قسمتی از یک شیء است که عقیده یا نظر خاصی در مورد آن بیان می‌شود. به‌عنوان مثال، به جمله زیر توجه کنید:

The story of the movie was weak and boring.

در این مثال، کلمه movie موجودیت، کلمه story که اسم است، ویژگی و کلمات weak و boring که صفت هستند احساس یا نظر بیننده را منعکس می‌کنند [۴۰].

از آن‌جا که اسناد را نمی‌توان به‌طور مستقیم پردازش نمود باید هر سند را به‌صورت برداری از وزن کلمات به شکل زیر نمایش داد که به آن، بسته لغات می‌گویند:

$$d_j=\{w_{1j},w_{2j},\ldots,w_{|T|j}\} \quad (۴)$$

در این رابطه، T مجموعه‌ای از کلمات (مرتبط با ویژگی‌ها) است که حداقل یک‌بار در حداقل یک سند آموزشی رخ می‌دهد و $0\le w_{kj}\le 1$ نشان می‌دهد که کلمه t_k چند بار در سند d_j اتفاق می‌افتد.

وزن کلمات را می‌توان به‌وسیله $tfidf$ نرمال به شکل زیر مشخص نمود:

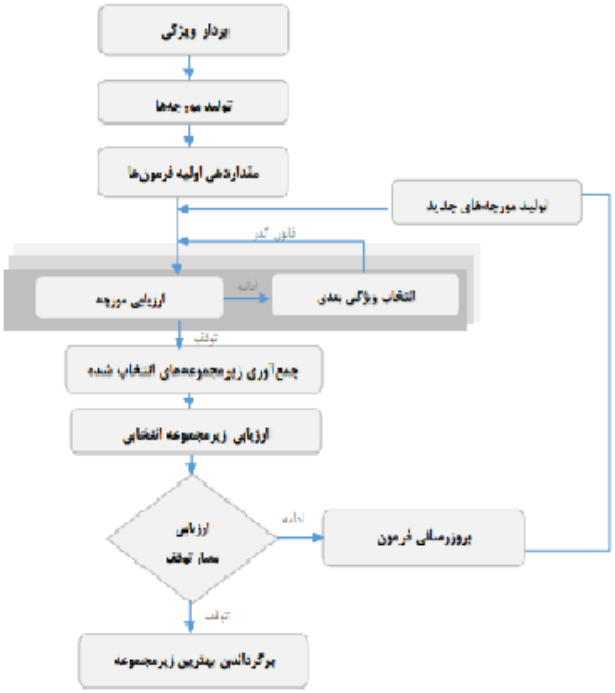
$$w_{kj}=\frac{tfidf(t_k,d_j)}{\sqrt{\sum_{s=1}^{|T|}(tfidf(t_s,d_j))^2}} \quad (۵)$$

که در آن:

$$tfidf(t_k,d_j)=tf(t_k,d_j).\log\frac{|Tr|}{\#_{Tr}(t_k)} \quad (۶)$$

در این رابطه، $tf(t_k,d_j)$ به تعداد وقوع ویژگی t_k در سند d_j، Tr به مجموعه آموزشی، $|Tr|$ به اندازه مجموعه آموزشی و $\#_{Tr}(t_k)$ به تعداد اسناد در Tr که ویژگی t_k در آن‌ها اتفاق می‌افتد دلالت می‌کنند [۴۱].

در شکل ۴ مراحل انتخاب ویژگی از طریق الگوریتم کلونی مورچگان را در زبان مبدأ مشاهده می‌نمایید:



شکل ۴: مراحل انتخاب ویژگی در زبان مبدا به‌کمک ACO [۴۲]

www.SID.ir

Archive of SID

جدول ۲: نمونه‌ای از روند تعیین ویژگی‌ها

عملیات	خروجی
عقیده	An intriguing suspense film with a shocking ending. The clues are obvious throughout the film, but the viewer is, nevertheless, surprised and possibly amazed.
تقسیم جملات	An intriguing suspense film with a shocking ending. The clues are obvious throughout the film, (but) the viewer is, nevertheless, surprised and possibly amazed.
نقش کلمات	An (Determiner) intriguing (Adjective) suspense (Noun) film (Noun) with (Preposition) a (Determiner) shocking (Adjective) ending (Noun). The (Determiner) clues (Noun) are (Verb) obvious (Adjective) throughout (Preposition) the (Determiner) film (Noun) (but) (Conjunction) the (Determiner) viewer (Noun) is (Verb), nevertheless (Adverb), surprised (Adjective) and (Conjunction) possibly (Adverb) amazed (Adjective).
انتخاب ویژگی‌ها	film, clues, viewer
جمع‌آوری احساسات	file (intriguing, shocking) clues (obvious) viewer (surprised, amazed)

پس از انتخاب ویژگی، از طریق مراحل زیر میزان قطبیت نظرات زبان مقصد را با توجه به قطبیت نظرات زبان مبدأ مشخص می‌نماییم:

۱- صفت‌ها، قیدها و افعالی که برای یک ویژگی در نظرات زبان مبدأ به کار رفته‌اند را برای آن ویژگی ثبت خواهیم نمود.

۲- از طریق یک مترجم ویژگی‌ها، صفت‌ها، قیدها و افعال به زبان مقصد ترجمه خواهند شد.

۳- برای تمام ویژگی‌ها، در نظرات زبان مقصد، تمام لغات معادل مربوط به ویژگی‌ها را جستجو خواهیم نمود.

۴- پس از پیدا کردن ویژگی یا موضوع نظر در زبان مقصد، صفت‌ها، قیدها و افعال مختص آن ویژگی را در نظر مورد بررسی، جستجو خواهیم کرد.

۵- نحوه قرار گرفتن صفت‌ها، قیدها و ... نسبت به ویژگی مورد بررسی مشخص‌کننده ساختار جملات در زبان مقصد هستند. ساختار جملات یک نظر را از طریق محل قرارگیری اجزا آن ثبت خواهیم نمود.

۶- جملاتی از زبان مبدأ که مشابه ساختار جملات در زبان مقصد هستند را مشخص می‌کنیم (مشابه از نظر ویژگی و ساختار دستوری).

۷- میانگین قطبیت جملات زبان مبدأ نشان‌دهنده قطبیت جملات زبان مقصد هستند.

در جدول ۲، نمونه‌ای از نظرات مخاطبین در مورد دامنه DVD از مجموعه دامنه‌های آمازون (به‌عنوان زبان مبدا) و نحوه انتخاب ویژگی‌ها آورده شده است. مشاهده می‌شود که این عقیده (در سطح سند)، به تعدادی نظر در سطح جمله قابل‌تقسیم است که در مجموع، عقیده مثبتی را نشان می‌دهند. بعد از تعیین نقش کلمات (POS)، ویژگی‌های موجود (در نقش Noun) را به کمک الگوریتم مورچگان مشخص می‌کنیم و سپس در مورد هر ویژگی، صفات، قیود و افعال نشان‌دهنده احساس (Sentiment) را جمع‌آوری می‌نماییم.

روش پیشنهادی به دنبال سه هدف عمده است:

• اغلب کاربردهای یادگیری انتقالی در عقیده‌کاوی، در سطح سند عمل می‌کنند، درحالی‌که سطح جمله یا ویژگی نتایج بهتر و دقیق‌تری ایجاد می‌کند.

• پشتیبانی از ویژگی چندقطبی با در نظر گرفتن صفات، قیدها و افعال مربوط به یک ویژگی، چنین عملی موجب می‌گردد که با توجه به نتایجی که از زبان مبدأ کسب شده است میزان دقت سنجش قطبیت نظرات زبان مقصد را نیز افزایش داد.

• جلوگیری از تداخل داده‌های وابسته به زبان مبدأ و داده‌های زبان مقصد از طریق یک‌سوسازی روند ترجمه از مبدأ به سمت مقصد و همچنین مشخص کردن سبدهای لغوی که اهداف جستجو را در نظرات زبان مقصد معین می‌نمایند.

۵- نتایج

برای بررسی فرضیه خود، ما از تعداد ۱۵۲۷ عقیده از زبان انگلیسی به‌عنوان زبان مبدأ و تعداد ۳۰۰ عقیده از زبان فارسی به‌عنوان زبان مقصد استفاده کرده‌ایم. این عقاید در مورد محصولات مختلف حوزه تکنولوژی ارتباطات خصوصا گوشی تلفن همراه، با بررسی ویژگی‌هایی همچون قضاوت در مورد دوربین، کیفیت سیستم‌عامل و ... است. عقاید انگلیسی با ۵ برچسب شامل نظر با قطبیت خیلی مثبت، مثبت، نظر بدون قطبیت یا خنثی، منفی و نظر با قطبیت خیلی منفی برچسب گذاری شده‌اند.

به‌کمک ماتریس درهم‌ریختگی (جدول ۳) می‌توان میزان پیش‌بینی درست و نادرست سیستم را مشخص نمود که موارد زیر در آن تعریف می‌شوند:

- مثبت واقعی (TP): مدل به‌درستی متعلق بودن عقیده را به یک کلاس پیش‌بینی کرده است،

www.SID.ir

Archive of SID

- منفی واقعی (FP): مدل به‌اشتباه متعلق بودن عقیده را به یک کلاس پیش‌بینی کرده است،
- مثبت کاذب (TN): مدل به‌درستی عدم تعلق عقیده را به یک کلاس پیش‌بینی کرده است،
- منفی کاذب (FN): مدل به‌اشتباه عدم تعلق عقیده را به یک کلاس پیش‌بینی کرده است.

جدول ۳: ماتریس درهم‌ریختگی

		پیش‌بینی‌شده	
		مثبت	منفی
واقعی	مثبت	TP	FN
	منفی	FP	TN

جدول ۴ ماتریس درهم‌ریختگی نتایج حاصل از الگوریتم پیشنهادی را نشان می‌دهد:

جدول ۴: ماتریس درهم‌ریختگی روش پیشنهادی

		پیش‌بینی‌شده	
		مثبت	منفی
واقعی	مثبت	۹۶	۲۹
	منفی	۴۳	۱۳۲

در این پژوهش، برای ارزیابی اعتبار تعیین قطبیت عقاید، از سه مقیاس استفاده شده است: جامعیت، مانعیت و دقت. دلیل استفاده از این معیارها این است که کاربران به طور متوسط تمایل به بازیابی مقادیر زیادی از موضوع‌های مربوط به هم دارند (که جامعیت بالایی را ایجاد می‌کند)؛ این در حالی است که مایلند به‌طور هم‌زمان بخش اعظمی از موضوع‌های غیرمرتبط (که مانعیت بالایی را ایجاد می‌کند) حذف شود.

- جامعیت (recall) به معنای توانایی سیستم در بازیابی مدارک مرتبط،

$$recall = \frac{TP}{TP + FN} \quad (۷)$$

- مانعیت (precision) به معنای توانایی سیستم در کنار گذاشتن مدارک نامرتبط

$$precision = \frac{TP}{TP + FP} \quad (۸)$$

- دقت (accuracy) به معنای مقادیر درست پیش‌بینی شده به تعداد کل نمونه‌ها است:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (۹)$$

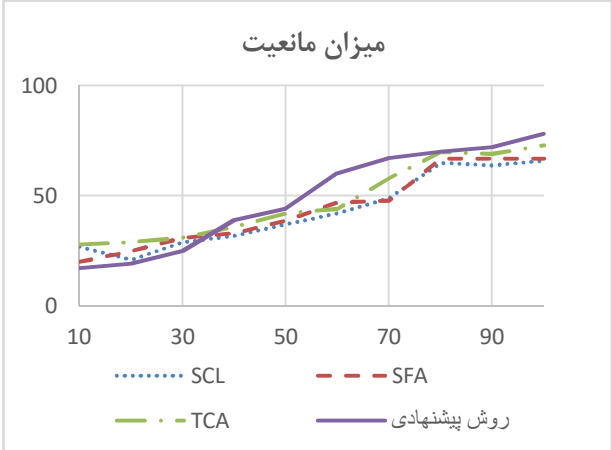
نتایج حاصل از آزمایش روش پیشنهادی و مقایسه آن با سه روش پایه و برجسته انتقال ویژگی SCL [45], SFA [44], TCA [43] روی داده‌های مذکور از نظر سه معیار جامعیت، مانعیت و دقت، در جدول شماره ۵ ارائه شده است. مشاهده می‌شود که روش پیشنهادی، بهبود قابل‌ملاحظه‌ای را از نظر سه معیار فوق نسبت به روش‌های پایه نشان می‌دهد. به‌عنوان‌مثال، در مورد پارامتر جامعیت، افزایش بازدهی قابل توجه ۹، ۵ و ۵ درصد نسبت به روشهای SCL، SFA و TCA حاصل شده است.

جدول ۵: مقایسه نتایج حاصل آزمایش روش پیشنهادی و مقایسه آن با سه روش انتقال ویژگی SCL,SFA,TCA

	جامعیت٪	مانعیت٪	دقت٪
SCL	۶۸	۶۵	۶۸
SFA	۷۲	۶۷	۶۹
TCA	۷۲	۷۰	۷۲
روش پیشنهادی	۷۷	۶۹	۷۶

قابل ذکر است که نتایج حاصل از جدول ۵ با به‌کارگیری ۸۰ درصد مقادیر موجود در زبان مبدأ برای یادگیری به دست آمده است. همچنین میزان تغییرات سه متغیر جامعیت، مانعیت و دقت اندازه‌گیری الگوریتم را در مقایسه با سه الگوریتم ذکر شده در جدول ۵ به صورت نمودار خطی در شکل‌های ۵، ۶ و ۷ رسم شده است.

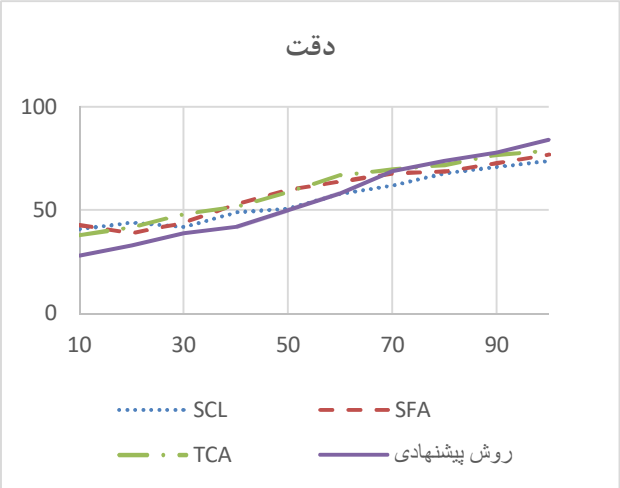
شکل ۵: مقایسه میزان جامعیت روش پیشنهادی با سه الگوریتم SCL, SFA, TCA با افزایش تعداد نظرات زبان مبدأ



شکل ۶: مقایسه میزان مانعیت روش پیشنهادی با سه الگوریتم SCL, SFA, TCA با افزایش تعداد نظرات زبان مبدأ

www.SID.ir

Archive of SID



شکل ۷: مقایسه میزان دقت روش پیشنهادی با سه الگوریتم SCL, SFA, TCA با افزایش تعداد نظرات زبان مبدأ

۶- نتیجه‌گیری

در پژوهش حاضر، به بررسی عقیده‌کاوی با استفاده از یادگیری انتقالی مبتنی بر انتقال ویژگی بین دو زبان انگلیسی به‌عنوان زبان مبدأ و زبان فارسی به‌عنوان زبان مقصد پرداختیم. با توجه به نتایج حاصل در بخش ۵ می‌توان گفت که روش پیشنهادی می‌تواند بدون دانستن قواعد موجود در زبان فارسی و تنها با استفاده از دامنه لغت‌نامه انگلیسی‌به فارسی، با دقت ۷۷ درصد میزان قطبیت نظرات را در زبان مقصد با استفاده از آموزش و برداشت گروه کلمات مرتبط و کشف ساختار دستوری از زبان مبدأ تشخیص دهد.

همچنین در آزمایش‌ها نشان دادیم با افزایش تعداد نظرات موجود در زبان مبدأ می‌توان نتایج بهتری از نظر سه معیار جامعیت، مانعیت و دقت کلی بازیابی قطبیت عقاید را مشاهده نمود.

مراجع

[۲] سعاد شریفات‌زاده و محمدعلی زارع چاهوکی، «یادگیری انتقالی با روش تلفیقی از انتقال نمونه و نمایش ویژگی برای پیش‌بینی نقص بین‌پروژه‌ای نرم‌افزار»، مجله مهندسی برق دانشگاه تبریز، جلد ۴۸، شماره ۱، صفحات ۱۱۲-۱۰۱، بهار ۱۳۹۷.

[1] D. M. Eler, D. Grosa, I. Pola, R. Garcia, R. Correia and J. Teixeira, "Analysis of document pre-processing effects in text and opinion mining," Information, vol. 9, no. 4, pp. 1-13, 2018.

[3] S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345-1359, 2010.

[4] K. Dashtipour, M. Gogate, A. Adeel, C. Ieracitano, H. Larijani and A. Hussain, "Exploiting deep learning for Persian sentiment analysis," in Proceeding of the 9th international conference on Advances in Brain Inspired Cognitive Systems (BICS 2018), pp. 597-605, Xi'an, China, 2018.

[5] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Foundations and Trends in Information Retrieval, vol. 2, no. 1-2, pp. 1-153, 2008.

[6] J. Barnes, R. Klinger and S. Schulte, "Bilingual sentiment embeddings: Joint protection of sentiment across languages," in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, pp. 2483-2493, Melbourne, Australia, 2018.

[7] H. Guo, H. Zhu, Z. Guo, X. Zhang, and Z. Su, "OpinionIt: A text mining system for cross-lingual opinion analysis," in Proceedings of the ACM conference on Information and Knowledge Management (CIKM-2010), pp. 1199-1208, Toronto, Canada, 2010.

[8] X. Wan, "Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis," in Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2008), pp. 553-561, Hawaii, USA, 2008.

[9] J. Brooke, M. Tofiloski, and M. Taboada, "Cross-linguistic sentiment analysis: From English to Spanish," in Proceedings of International Conference RANLP, pp. 50-54, Borovets, Bulgaria, 2009.

[10] X. Wan, "Co-training for cross-lingual sentiment classification," in Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP (ACL-IJCNLP-2009), pp. 235-243, Suntec, Singapore, 2009.

[11] B. Wei, and C. Pal, "Cross-lingual adaptation: An experiment on sentiment classifications," in Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 258–262, Uppsala, Sweden, 2010.

[12] K. Duh, A. Fujino, and M. Nagata, "Is machine translation ripe for cross-lingual sentiment classification?," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics:shortpapers (ACL-2011), pp. 429-433, Oregon, USA, 2011.

[13] C. Banea, R. Mihalcea, and J. Wiebe. "Multilingual subjectivity: are more languages better?," In Proceedings of the International Conference on Computational Linguistics (COLING-2010), pp. 28-36, Beijing, China, 2010.

[14] M. Bautin, L. Vijayarenu, and S. Skiena, "International sentiment analysis for news and blogs," in Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM-2008), pp. 19-26, Washington, U.S.A, 2008.

[15] B. Lu, C. Tan, C. Cardie, and B. K. Tsou, "Joint bilingual sentiment classification with unlabeled parallel corpora," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL-2011), pp. 320-330, Oregon, USA, 2011.

[16] M. S. Rasooli, N. Farra, A. Radeva, T. Yu, and K. McKeown, "Cross-lingual sentiment transfer with limited resources," Machine Translation, vol. 32, no. 1-2, pp. 143-165, 2018.

[17] H. K. Aldayel and A. M. Azmi, "Arabic tweets sentiment analysis - a hybrid scheme," Journal of Information Science, vol. 42, no. 6, pp. 782-797, 2015.

[18] G. Vinodhini and R. Chandrasekaran, "Sentiment analysis and opinion mining: A survey," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 2, no. 6, pp. 282–292., 2012.

[19] I. Habernal, T. Pt´a˘cek and J. Steinberg, "Supervised sentiment analysis in Czech social media," Information Processing and Management, vol. 50, no. 5, pp. 693–707, 2014.

[20] H. Ghorbel and D. Jacot, *Advances in distributed agent-based retrieval tools*, Springer Berlin Heidelberg, 2011.

[21] T. Scholz and S. Conrad, "Linguistic sentiment features for newspaper opinion mining," 18th International Conference on Applications of Natural Language to Information Systems (NLDB 2013), pp. 272–277, Salford, UK, 2013.

[22] N. Medagoda, S. Shanmuganathan and J. Whalley, "A comparative analysis of opinion mining and sentiment classification in non-English languages," in Proceedings of International Conference on Advances in ICT for Emerging Regions (ICTER 2013), pp. 144-148, Colombo, Sri Lanka, 2013.

[23] F. Neri, C. Aliprandi, F. Capeci, M. Cuadros and T. By, "Sentiment analysis on social media," in Proceedings of the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 951–958, Istanbul, Turkey, 2012.

[24] Y. Arakawa, A., Kameda, A., Aizawa and T. Suzuki, "Adding Twitter-specific features to stylistic features for classifying tweets by user type and number of retweets," Journal of the Association

www.SID.ir

Archive of SID

for Information Science and Technology, vol. 65, no. 7, pp. 1416–1423, 2014.

[25] N. Yussupova, D. Bogdanova, and M. Boyko, "Applying of sentiment analysis for texts in Russian based on machine learning approach," in Proceedings of 2nd International Conference on Advances in Information Mining and Management, pp. 8-14, Venice, Italy, 2012.

[26] D. Vilares, M. A. Alonso and C. G´omez-Rodr´ıguez, "A syntactic approach for opinion mining on Spanish reviews," Natural Language Engineering, vol. 21, no. 1, pp. 139-163, 2015.

[27] P. Inrak and S. Sinthupinyo, "Applying latent semantic analysis to classify emotions in Thai text," 2nd International Conference on Computer Engineering and Technology (ICCET 2010), pp. 450–454, Chengdu, China, 2010.

[28] M. Dorigo and C. Blum, "Ant colony optimization theory: A survey," Theoretical Computer Science, vol. 344, no. 2-3, pp. 243-278, 2005.

[۲۹] صمد نجاتیان، روح‌اله امیدوار، حمید پروین، وحیده رضایی و میلاد یثربی، «یک الگوریتم جدید: الگوریتم کلونی موش‌های وحشی»، مجله مهندسی برق دانشگاه تبریز، جلد ۴۹، شماره ۱، صفحات ۴۳۷-۴۲۵، بهار ۱۳۹۸.

[30] S. R. Ahmad, A. Abu Bakr, and M. R. Yaakub, "Ant colony optimization for text feature selection in sentiment analysis," Intelligent Data Analysis, vol. 23, no. 1, pp.133-158, 2019.

[31] M. Shams, A. Shakery and H. Faili, "A non-parametric LDA-based induction method for sentiment analysis," 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012), pp. 216-221, 2012.

[32] A. Bagheri and M. Sarae, "Persian sentiment analyzer: A framework based on a novel feature selection mthod," International Journal of Artificial Intelligence, vol. 12, no. 2, pp. 115-129, 2014.

[33] M. Basiri, A. Naghsh-Nilchi and N. Ghasem-Aghaee, "A framework for sentiment analysis in Persian," Open Transactions on Information Processing, vol. 1, no. 3, pp. 1-14, 2014.

[34] S. Alimardani and A. Aghaei, "Opinion mining in Persian language using supervised algorithms," Journal of Information Systems and Telecommunication, vol. 3, no. 3, pp. 135-141, 2015.

[35] F. Amiri. S. Scerri and M. Khodashahi, "Lexicon-based sentiment analysis for Persian text," in Proceedings of the International Conference on Recent Advances in Natural Language Processing, pp. 9-16, , Hissar, Bulgaria, 2015.

[36] E. Golpar-Rabooki, S. Zarghamifar, and J. Rezaeenour, "Feature extraction in opinion mining through Persian reviews," Journal of AI and Data Mining, vol. 3, no. 2, pp. 169-179, 2015.

[37] E. Asgarian, M. Kahani and S. Sharifi, "The Impact of Sentiment Features on the Sentiment Polarity Classification in Persian Reviews," Cognitive Computation, vol. 10, no. 1, pp. 117-135, 2018.

[38] S. Nemati, M. Basiri, N. Ghasem-Aghaee and M. Hosseinzadeh Aghdam, "Ant colony optimization for text feature selection in sentiment analysis," Expert Systems with Applications, vol. 36, no. 10, pp.12086-12094, 2009.

[39] K. Denecke, "Are SentiWordNet scores suited for multi-domain sentiment classification?," Fourth International Conference on Digital Information Management (ICDIM 2009), pp. 1–6, Michigan, USA, 2009.

[40] V. A. Kharde and S. S. Sonawane, "Sentiment analysis of twitter data: A survey of techniques," International Journal of Computer Applications, vol. 139, no. 11, pp. 5-15, 2016.

[41] F. Debole and F. Sebastiani, "Supervised term weighting forautomated text categorization," in Proceedings of the 2003 ACM Symposium on Applied Computing, pp. 784-788, New York, USA, 2003.

[42] M. Hosseinzadeh Aghdam, N. Ghasem-Aghaee and M. Basiri, "Text feature selection using ant colony optimization," Expert Systems with Applications, vol. 36, no. 3, pp.6843-6853, 2009.

[43] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," IEEE Transactions on Neural Networks, vol. 22, no. 2, pp. 199–210, 2011.

[44] S. J. Pan, X. Ni, J. T. Sun, Q. Yang, and Z. Chen, "Cross-domain sentiment classification via spectral feature alignment," in Proceedings of the 19th International Conference on World Wide Web (WWW 2010), pp. 751–760, North Carolina, USA, 2010.

[45] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification," in Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (ACL), pp. 440–447, Prague, Czech Republic, 2007.

زیرنویس ها

1 Dorigo
2 Multiagent
3 Traveling Salesman Problem (TSP)
4 Pheromone
5 Subset Problem
6 Bag of Words
7 Dempster-Shafer
8 Co-reference Resolution
9 Natural Language Processing
10 Persian Sentiment WordNet
11 Corpus
12 Candidate

www.SID.ir