



آنالیز احساسات در موضوعات فارسی

مبتنی بر تلفیق قوانین

مینا کشاورزی فراشاه^۱، محمدعلی زارع چاهوکی^۲

^۱ دانشجوی ارشد گرایش نرم‌افزار، گروه مهندسی کامپیوتر، پردیس فنی، دانشگاه یزد - m.keshavarzi@stu.yazd.ac.ir

^۲ استادیار و عضو هیات علمی، گروه مهندسی کامپیوتر، پردیس فنی، دانشگاه یزد - chahooki@yazd.ac.ir*

چکیده

با ظهور و رشد سریع رسانه‌های اجتماعی، این امکان برای کاربران فراهم شده تا افکار و نظرات خود را به سهولت در مورد موضوعات مختلف، رویدادهای سیاسی و اجتماعی بیان کنند. درک افکار عمومی می‌تواند در بهبود تصمیم‌گیری تاثیر بسزایی داشته باشد. حجم زیاد و فقدان ساختار مناسب برای متن نظرات ارائه شده در این رسانه‌ها، استفاده از دانش پنهان درون آن‌ها را دشوار نموده است. عقیده‌کاوی یا آنالیز احساسات راه‌حلی مناسب برای تحلیل نظرات به صورت خودکار می‌باشد. در این پژوهش روشی برای تحلیل نظرات افراد در مورد موضوعات فارسی تلگرام ارائه می‌شود. در این راستا سعی شده با استفاده از روش‌های مبتنی بر قانون و با بهره‌گیری از کیسه کلمات به تحلیل نظرات کاربران بپردازیم. در ادامه به ارزیابی روش، با پیاده‌سازی آن روی مجموعه داده جمع‌آوری شده از نظرات به زبان فارسی پرداخته شده است. نتایج آزمایش‌های تجربی نشان می‌دهد که رویکرد استخراج قوانین با بهره‌گیری از کیسه کلمات کارایی مناسبی در عقیده‌کاوی به زبان فارسی دارد و همچنین تلفیق قوانین سبب بهبود دقت می‌شود.

کلمات کلیدی: عقیده‌کاوی - آنالیز احساسات - تلفیق قوانین - کیسه کلمات - موضوعات فارسی تلگرام

Sentiment Analysis in Persian subjects by fusion of rules

Mina Keshavarzi farashah Mohammad Ali Zare Chahooki

Abstract

With the advent and rapid growth of social media, it is possible for users to express their thoughts and opinions easily on various topics, political and social events. Understanding public opinion can have a significant impact on decision-making. The large volume and the lack of proper structure for the text of the comments presented in these media makes it difficult to use the hidden knowledge within them. Opinion Mining or Sentiment Analysis is the ideal way to automatically analyze comments. In this research, a method is presented to analyze the views of individuals on the Persian telegram subjects. In this regard, we tried to analyze the opinions of users using the rule-based methods and using the bag of words. Subsequently, the method was evaluated by implementing it on the dataset aggregated from comments in Persian. The results of empirical experiments show that the rule extraction approach using the bagofwords has a good performance in Persian, and also the fusion of rules improves accuracy.

تاریخچه مقاله:

تاریخ ارسال: ۱۳۹۷/۳/۲۰

تاریخ اصلاحات: ۱۳۹۷/۵/۱

تاریخ پذیرش: ۹۷/۵/۵

تاریخ انتشار: ۹۷/۵/۱۵

Keywords:

Opinion Mining
Sentiment Analysis
fusion rules
Bag Of Words
Persian Telegram subjects

م. کشاورزی فراشاه، م.ع. زارع چاهوکی، آنالیز احساسات در موضوعات فارسی مبتنی بر تلفیق قوانین، دوفصلنامه محاسبات و سامانه‌های توزیع شده، سال اول، شماره اول، ص ۳۱-۴۶، سال انتشار ۱۳۹۷.

روش ارجاع به مقاله:

* نویسنده عهده دار مکاتبات: محمدعلی زارع چاهوکی



۱- مقدمه

با رشد سریع شبکه‌های اجتماعی^۱ و وبلاگ‌ها^۲، سرعت استفاده از قابلیت‌های این رسانه‌های اجتماعی توسط کاربران آنلاین افزایش یافته است. کاربران نظرات و تجربیات خود را در مورد یک محصول، سیاست و یا یک رویداد خاص به اشتراک می‌گذارند [۱]. شناسایی و کشف دانش موجود در متن نظرات یکی از موضوعات مطالعاتی مهم محسوب می‌شود. با توجه به اینکه این کشف با مشکلاتی از قبیل حجم بالا و عدم وجود ساختار مناسب در متن نظرات مواجه است، باید روش‌هایی ارائه شود تا بتواند این دانش را استخراج کند. بنابراین عقیده‌کاوی^۳ راه‌حلی مناسب برای تحلیل و کشف دانش موجود در نظرات به صورت خودکار می‌باشد [۲].

تکنیک‌های عقیده‌کاوی، نظرات افراد را در مورد موجودیت‌ها، حوادث و موضوعات استخراج می‌کنند و اطلاعات ارزشمندی را از داده‌ها فراهم می‌کنند. به طور کلی عقیده‌کاوی را می‌توان در سه دسته (۱) توسعه منابع زبانی^۴، (۲) طبقه‌بندی احساسات^۵ و (۳) خلاصه‌سازی نظرات^۶ تقسیم‌بندی کرد [۳].

عقیده‌کاوی زمینه تحقیقاتی رو به رشدی می‌باشد که با علوم مختلفی از قبیل بازیابی اطلاعات^۷،

داده‌کاوی^۸ و تکنیک‌های پردازش زبان طبیعی^۹ بسیار مرتبط می‌باشد. هدف رایج از عقیده‌کاوی، تحلیل و بررسی نظرات افراد و سپس تشخیص قطبیت^{۱۰} موجود در آن نظرات (مثبت^{۱۱}، منفی^{۱۲} و یا خنثی^{۱۳}) به صورت خودکار می‌باشد. بنابراین به عقیده‌کاوی، آنالیز احساسات^{۱۴} نیز گفته می‌شود [۴].

برای فرایند عقیده‌کاوی، روش‌های بسیاری وجود دارد که می‌توان آن‌ها را به سه دسته (۱) رویکردهای مبتنی بر واژگان احساسی^{۱۵}، (۲) رویکردهای مبتنی بر یادگیری ماشین^{۱۶} و (۳) رویکردهای ترکیبی^{۱۷} تقسیم کرد [۵،۶]. البته برای عقیده‌کاوی در دادگان حجیم^{۱۸} از روش‌های مبتنی بر گراف^{۱۹} استفاده می‌شود [۷].

از نگاهی دیگر، روش‌های عقیده‌کاوی به سه دسته کلی سطح متن^{۲۰}، سطح جمله^{۲۱} و سطح مؤلفه^{۲۲} تقسیم می‌شوند. مقصود از عقیده‌کاوی در سطح متن تعیین گرایش معنایی کل متن و در سطح جمله، تعیین احساس یک جمله به صورت مستقل است. آنالیز در سطح مؤلفه که هم از طرف جوامع علمی و هم صنعت مورد توجه است، با هدف کشف احساس درباره یک موجودیت و یا یک مؤلفه خاص از آن می‌باشد [۸،۹،۱۰].

اگر چه مطالعات اولیه در عقیده‌کاوی در متون انگلیسی^{۲۳} متمرکز شده است، برخی از کارهای اخیر

¹² Negative

¹³ Neutral

¹⁴ Sentiment Analysis

¹⁵ lexicon-based

¹⁶ Machine Learning

¹⁷ Hybrid

¹⁸ Big Data

¹⁹ Graph-based

²⁰ Document level

²¹ Sentence level

²² Aspect level

²³ English

¹ Social Networks

² weblogs

³ Opinion mining

⁴ linguistic resource development

⁵ sentiment classification

⁶ opinion summarization

⁷ Information Retrieval

⁸ Data Mining

⁹ Natural Language Processing (NLP)

¹⁰ Polarity

¹¹ Positive



در این پژوهش، دادگانی^{۲۸} شامل پیام‌های منتشر شده در گروه‌های تلگرام فارسی در موضوعات سیاسی، اجتماعی، ورزشی و فرهنگی ساخته شده است. این دادگان می‌تواند برای عقیده‌کاوی و یا تحلیل نظرات در دیگر پژوهش‌ها نیز مورد استفاده قرار گیرد.

در این پژوهش عقیده‌کاوی مبتنی بر قانون با استخراج قوانین به صورت خودکار صورت می‌پذیرد. لذا در سه مرحله قوانینی با یک، دو و سه کلمه استخراج می‌شوند و در نهایت تصمیم‌گیری با کمیته‌ای از این سه سطح از قانون انجام می‌شود. نتایج آزمایش‌های تجربی نشان می‌دهد که رویکرد استخراج قوانین با بهره‌گیری از کیسه کلمات^{۲۹} کارایی مناسبی در عقیده‌کاوی زبان فارسی دارد و همچنین تلفیق^{۳۰} قوانین در سه سطح سبب افزایش دقت می‌شود.

۲- روش پیشنهادی

مراحل روش پیشنهادی این پژوهش، در شکل (۱) نشان داده شده است. در مرحله تهیه دادگان آموزش و آزمون، مجموعه داده مورد نیاز جهت عقیده‌کاوی به زبان فارسی ساخته می‌شوند. قبل از انجام مراحل اصلی بروی دادگان، پیش‌پردازشی^{۳۱} بروی آنها انجام می‌شود. روش پیشنهادی مبتنی بر استخراج قوانین با استفاده از کلمه‌های تکی، دوتایی و سه‌تایی است، بنابراین ساخت کیسه کلمه‌ها یکی از مراحل مهم به شمار می‌رود. در مرحله بعد با توجه به کیسه کلمات ساخته شده قوانین استخراج می‌شوند. با بهره‌گیری از قوانین قطبیت

بروی زبان‌های دیگر نیز انجام شده است [۱۱]. با این حال، کارهای کمی برای عقیده‌کاوی به زبان فارسی^{۲۴} وجود دارد [۱۱]. با توجه به اینکه بیشتر پژوهش‌های انجام شده در این حوزه بروی شبکه‌های اجتماعی توییتر^{۲۵} و فیس‌بوک^{۲۶} بوده و بروی دادگان فارسی کار زیادی انجام نشده است، شبکه اجتماعی تلگرام^{۲۷} می‌تواند انتخاب مناسبی برای شروع تحقیقی موثر در این حوزه باشد.

شبکه اجتماعی تلگرام با ردوبدل کردن روزانه ۱۵ میلیارد پیام از پرطرفدارترین شبکه‌های اجتماعی در ایران محسوب می‌شود [۱۲]. مردم ایران در محیط تلگرام بیش از ۱۷۰ هزار شبکه تلگرامی به راه انداخته اند، که در این میان ۱۱ هزار کانال بیش از ۵۰۰۰ نفر عضو دارد. هر کاربر ایرانی بطور میانگین در ۱۰ کانال عضویت دارد و روزانه بیش از ۱۰۰ مطلب را از تلگرام مطالعه می‌کند [۱۲].

گسترش تحقیقات در این حوزه در سال‌های اخیر دلایل متعددی دارد. اولین دلیل، کاربردهای فراوان عقیده‌کاوی در زمینه‌های مختلف می‌باشد. از دیگر دلایل می‌توان به وجود چالش تحقیقاتی به خصوص در زبان فارسی با توجه به پیچیدگی‌های خاص این زبان اشاره کرد [۹]. از مهم‌ترین کاربردهای عقیده‌کاوی می‌توان به دنبال کردن عقاید مردم توسط سیاستمداران، آگاهی تولیدکنندگان از سطح رضایت مشتریان و پیش‌بینی تغییرات بازار با توجه به نظرات افراد اشاره کرد.

²⁸ Dataset

²⁹ Bag Of Word (BOW)

³⁰ fusion

³¹ Preprocess

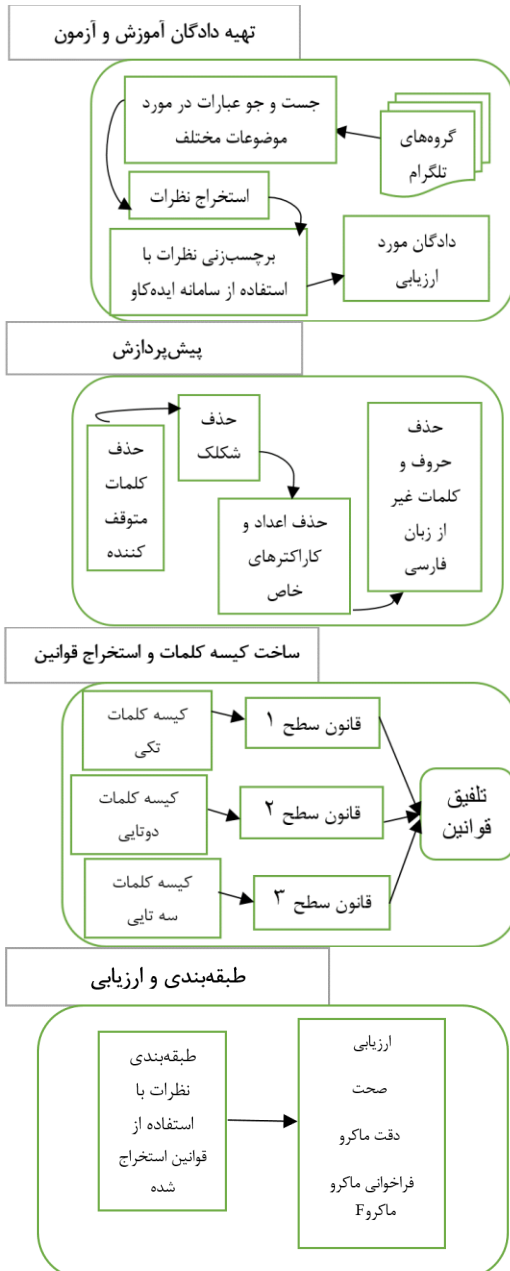
²⁴ Persian

²⁵ Twitter

²⁶ Facebook

²⁷ Telegram

۴- اگر در مورد موجودیتی هم نظر مثبت و هم نظر منفی بیان شده است، نظری که شدت بیشتری دارد، قطبیت پیام را تعیین می کند. اگر شدت نظرات یکسان باشد، قطبیت متن خنثی می باشد.



(شکل-۱): دیاگرام روش پیشنهادی

(Figure-1): Diagram of the proposed method

دادگان به صورت خودکار تعیین می شود. در آخرین مرحله نتایج مورد ارزیابی قرار می گیرند.

۲-۱- تهیه دادگان آموزش و آزمون

روش پیشنهادی این پژوهش، نیازمند مجموعه دادگان جمع آوری شده در شبکه های اجتماعی به زبان فارسی می باشد. برای رفع این نیازمندی اولین مرحله آماده سازی این نوع دادگان می باشد. دادگان مورد استفاده در این پژوهش، شامل ۱۰۰۰۰ پیام منتشر شده در شبکه اجتماعی تلگرام می باشد. با توجه به این که هدف عقیده کاوی در سطح جمله می باشد، فقط پیام های با طول بین ۲۰ تا ۲۰۰ کاراکتر در دادگان وجود دارد. برای حصول اطمینان از صحت برچسب ها، هر پیام توسط سه نفر برچسب گذاری شده است. به منظور همگون سازی نحوه ی برچسب گذاری توسط هر فرد، قوانینی تهیه شده است. این قوانین کمک می کند تا ابهام کمتری در برچسب گذاری پیام ها ایجاد شود. این قوانین شامل موارد زیر می باشد:

۱- جهت فکری افراد در فرایند برچسب گذاری پیام ها دخالتی ندارد. هر آن چه که متن پیام به مخاطب القا می کند، بدون پیش داوری در نظر گرفته می شود.

۲- در نقل قول ها، گوینده ملاک برچسب گذاری نیست. موجودیت اصلی، آن چیز یا کسی است که درباره ی آن صحبت شده است. در صورتی که متن دارای بیش از یک موجودیت باشد، ملاک برچسب گذاری اولین موجودیت می باشد.

۳- شکلک ها و علائم نگارشی به کار رفته در متن اهمیت زیادی در تشخیص قطبیت متن دارند.



۳. حروف و کلمات غیر از زبان فارسی: تمامی حروف و کلمات غیر از زبان فارسی مانند حروف و کلمات انگلیسی باید حذف شود.

۴. شکلک^{۳۳}: ساده ترین روش تشخیص حالت یک نویسنده، مشاهده ی شکلک هایی است که فرد در متن به کار برده است. با این شکلک های بسیار ساده می توان مفاهیم احساسی پیچیده ای را انتقال داد. با وجود اینکه شکلک ها در عقیده کاوی بسیار موثر هستند، اما به دلایل زیر نادیده گرفته می شوند:

- استفاده نادرست کاربر: همه افراد از شکلک ها برای بیان احساسات خود استفاده نمی کنند. این امکان وجود دارد که افراد برحسب عادت، در پایان متن شکلک مثبت به کار ببرند.
- کاربرد بسیار کم شکلک ها: در متن های جدی شکلک ها کاربر بسیار کمی دارند. به خصوص کاربران هیجان زده یا بسیار عصبانی ممکن است اصلاً از شکلک ها استفاده نکنند. پژوهش [۱۴] نشان داده که میزان استفاده از شکلک ها در محیط های رسمی کمتر از ۱۰٪ می باشد.

جهت سهولت برچسب گذاری سامانه ای به نام ایده کاو [۱۳] طراحی شده که عملیات برچسب گذاری با استفاده از آن انجام می شود. حدود ۶۰ عبارت با مضامین سیاسی، فرهنگی و اجتماعی به برچسب گذاران داده شد. آن ها با جستجوی این عبارات، پیام های بازبایی شده را برچسب گذاری می کردند.

۲-۲- پیش پردازش

در روش پیشنهادی ابتدا باید متن ورودی پیش پردازش شود. در روند اجرای روش پیشنهادی، عملیاتی از قبیل حذف کلمات متوقف کننده و داده های اضافی برای آماده سازی متن مورد نیاز است. نمونه ای از داده های اضافی شامل موارد زیر می باشد:

۱. کلمه های متوقف کننده^{۳۲}: این کلمات هیچ نقشی در تعیین احساسات ندارند. نمونه ای از این کلمات "است"، "باید"، "از"، "تا" و غیره می باشد. در این پژوهش لیستی از کلمه های متوقف کننده مورد استفاده در زبان فارسی ایجاد شده است.
۲. اعداد و کاراکترهای خاص: کاراکترهایی مانند @، #، \$، % و اعداد تاثیری در تعیین احساسات ندارند. وجود چنین کاراکترها و اعداد در متون گیج کننده می باشد.

³² Stop Words

³³ Emoticon



۲-۳- ساخت کیسه کلمات

(جدول-۲): کلمات تکی و تعداد تکرار آن در هر کلاس

(Table-2): Single words and its number in each class

کلمات تکی	تعداد تکرار در جملات منفی	تعداد تکرار در جملات مثبت	تعداد تکرار در جملات نامشخص	مجموع
'حسن'	۶۸	۱۵۴	۵۴۰	۷۶۲
'جوهرچی'	۶	۱۱۳	۴۲۷	۵۴۶
'درگذشت'	۱۰	۴۳	۱۶۳	۲۱۶
'روحش'	۰	۸۱	۱۱	۹۲
'شاد'	۲۵	۱۰۷	۶۱	۱۹۳
'پادش'	۶	۴۹	۹	۶۴
'گرامی'	۱۲	۷۵	۵۶	۱۴۳

با توجه به تعداد زیاد کلمات در دادگان، ترکیب کلمات دوتایی و سه تایی فراوانی بدست می آید. پردازش این ترکیبات بسیار زمان بر و طولانی می باشد، برای رفع این مشکل باید تعداد کلمات در دادگان کاهش یابد. با در نظرگیری یک بازه مشخص از فرکانس یا تعداد تکرار کلمات در مجموع جملات می توان تعداد آن ها را کاهش داد.

در این پژوهش برای استخراج ترکیبات دوتایی و سه تایی، کلمات دارای تعداد تکرار بالاتر از ۵۰۰ از دادگان حذف شده اند. بنابراین دادگان جدید با تعداد کلمات کمتر ایجاد می شود.

برای ساخت کیسه کلمات دوتایی، تمامی ترکیبات دوتایی غیر تکراری از کلمات موجود در جملات دادگان جدید استخراج می شود. سپس با توجه به برچسب جملات، تعداد رخداد کلمات دوتایی در هر یک از کلاس های مثبت، منفی و یا نامشخص محاسبه

یکی از مهم ترین مراحل روش پیشنهادی این پژوهش، ساخت کیسه کلمه ها با ترکیب های مختلف کلمه ها می باشد. هدف از این مرحله، ایجاد کردن سه نوع کیسه کلمات تکی، دوتایی و سه تایی می باشد. سپس با در نظرگرفتن هر کدام از این کیسه کلمات قوانینی استخراج می شود. با بهره گیری از قوانین نظرات طبقه بندی شده و معیارهای ارزیابی برای مدل های ایجاد شده سنجیده می شود.

برای ساخت کیسه کلمات از مجموعه داده های آموزشی استفاده می شود. با توجه به اینکه برخی از جملات دادگان مورد استفاده، تنها از یک شکلک و یا ترکیب چند شکلک استفاده شده اند، بعد از مرحله پیش پردازش این جملات تهی می شوند. قبل از ساخت کیسه کلمات تکی باید این جملات تهی شناسایی و حذف شوند.

در کیسه کلمات تکی، جملات به کلمات یکتا شکسته می شوند. با توجه به برچسب جملات، تعداد رخداد کلمات در هر یک از کلاس های مثبت، منفی و خنثی محاسبه می شود. جدول (۱) نمونه جمله و جدول (۲) کلمات تکی از آن را که بخشی از کیسه کلمات کلی است، نشان داده شده است.

(جدول-۱): نمونه جمله از دادگان آموزشی به همراه برچسب آن

(Table-1): Sample sentence of dataset with its label

برچسب	جمله
مثبت	حسن جوهرچی درگذشت روحش شاد پادش گرامی



ترکیب‌های سه تایی کلمات	تعداد تکرار در جملات منفی	تعداد تکرار در جملات مثبت	تعداد تکرار در جملات نامشخص	مجموع
'شاد گرامی یادش'	۰	۳۶	۳	۳۹
'روحش شاد گرامی'	۰	۲۵	۳	۳۸
'درگذشت روحش گرامی'	۰	۱۰	۱	۱۱
'درگذشت گرامی یادش'	۰	۱۳	۱	۱۴
'روحش گرامی یادش'	۰	۳۶	۳	۳۹
'روحش شاد یادش'	۰	۳۷	۴	۴۱
'درگذشت روحش شاد'	۰	۳۵	۳	۳۸
'درگذشت شاد یادش'	۰	۱۱	۱	۱۲
'درگذشت شاد گرامی'	۰	۱۰	۱	۱۱
'درگذشت روحش یادش'	۰	۱۱	۱	۱۲

می‌شود. ترکیب دوتایی از کلمات جمله نمونه در جدول (۳) نشان داده شده است.

(جدول-۳): ترکیب دوتایی کلمات و تعداد تکرار آن در هر کلاس

(Table-3): The dual combination of words and the number of repetitions in each class

ترکیب‌های دوتایی کلمات	تعداد تکرار در جملات منفی	تعداد تکرار در جملات مثبت	تعداد تکرار در جملات نامشخص	مجموع
'روحش گرامی'	۰	۳۶	۳	۳۹
'شاد گرامی'	۰	۳۶	۳	۳۹
'شاد یادش'	۰	۳۸	۴	۴۲
'درگذشت گرامی'	۰	۱۳	۲	۱۵
'روحش شاد'	۰	۷۹	۱۱	۹۰
'درگذشت یادش'	۰	۱۴	۱	۱۵
'روحش یادش'	۰	۲۸	۴	۴۲
'درگذشت شاد'	۰	۲۵	۴	۳۹
'گرامی یادش'	۰	۴۴	۴	۴۸
'درگذشت روحش'	۰	۳۵	۳	۳۸

۲-۴- استخراج قانون

هدف از این پژوهش، مقایسه روش‌های مبتنی بر قانون استخراج شده از ترکیب‌های مختلف کلمات و تاثیر این ترکیب‌ها در فرایند عقیده‌کاوی می‌باشد. در این راستا می‌توان سه سطح قانون استخراج کرد. برای استخراج قوانین از کیسه کلمه‌ها استفاده می‌شود.

جمله‌های داده آموزشی به صورت کلمه‌های تکی، دوتایی و سه تایی تفکیک می‌شوند. با توجه به کیسه کلمه‌ها در مورد مثبت، منفی و یا خنثی بودن

ساخت کیسه کلمات سه تایی مشابه روش ساخت کیسه کلمه‌های دوتایی می‌باشد، با این تفاوت که تمامی ترکیبات سه تایی غیر تکراری از کلمات استخراج می‌شود. ترکیب سه تایی کلمات مربوط به جمله نمونه در جدول (۴) نشان داده شده است.

(جدول-۴): ترکیب سه تایی کلمات و تعداد تکرار آن در هر کلاس

(Table-4): A trivial combination of words and the number of repetitions in each class



که اگر تعداد رخداد هر ترکیب تکی، دوتایی و یا سه تایی از حدی کمتر باشد، از لیست ترکیب های جمله مربوطه حذف می گردد. تغییر دادن مقدار α می تواند معیارهای ارزیابی را تحت تاثیر قرار دهد. برای بدست آوردن α بهینه از اعتبارسنجی دستی و تغییر مقادیر این پارامتر بروی دادگان آموزش استفاده می شود.

در نهایت تصمیم گیری با کمیته ای از این سه سطح از قانون انجام می شود. در اینجا نیز معیار تصمیم گیری رای گیری اکثریت می باشد.

۲,۴,۱ قانون مبتنی بر کلمه های تکی

برای استخراج قوانین مبتنی بر کلمه های تکی از کیسه کلمه های تکی استفاده می شود. جمله های داده آموزشی به صورت کلمه های تکی تفکیک می شود. هر تک کلمه در کیسه کلمه ها جست و جو شده و با توجه به مقادیر محاسبه شده در مورد مثبت، منفی و یا خنثی بودن آن تک کلمه تصمیم گیری می شود.

در نتیجه برای هر جمله تعدادی تک کلمه با برچسب مشخص وجود دارد. برچسب نهایی در هر دو مدل با استفاده از تکنیک رای گیری اکثریت تعیین می - گردد.

۲,۴,۲ قانون مبتنی بر کلمات دوتایی

برای استخراج قوانین مبتنی بر کلمه های دوتایی از کیسه کلمه های دوتایی استفاده می شود. تمامی ترکیب های دوتایی از کلمه های موجود در جمله های داده آموزشی استخراج می شود. برای هر جمله به صورت مجزا ترکیب های دوتایی مربوطه مورد پردازش

ترکیب های مختلف کلمه تصمیم گیری می شود. معیار تصمیم گیری وجود بیشترین مقدار در تعداد تکرار و یا مجموع احتمال قطبیت کلمه مربوطه می باشد. در صورت مساوی بودن این معیار، برچسب خنثی برای آن ترکیب در نظر گرفته می شود. با توجه به معیار تصمیم گیری می توان دو مدل ایجاد کرد:

۱. مدل مبتنی بر تکرار قطبیت کلمه (RWP)^{۳۴}

۲. مدل مبتنی بر مجموع احتمال قطبیت کلمه (TPWP)^{۳۵}

در نتیجه برای هر جمله تعدادی تک کلمه یا دو کلمه ای و یا سه کلمه ای با برچسب مشخص وجود دارد. برای رسیدن به برچسب واحد برای جمله مربوطه، از تکنیک رای گیری اکثریت^{۳۶} استفاده می شود. برای مدل RWP، برچسب با بیشترین تکرار به عنوان خروجی و برچسب نهایی جمله محسوب می شود. در مدل TPWP، برچسب با بیشترین احتمال به عنوان برچسب پیش بینی شده برای آن جمله در نظر گرفته می شود. احتمال هر برچسب از نسبت تعداد تکرار برچسب مربوطه به کل تعداد تکرار آن در مجموعه دادگان محاسبه می شود.

با توجه به اینکه تعداد ترکیب های تکی، دوتایی و سه تایی کلمه ها زیاد است، پردازش روی آنها بسیار زمان بر و طولانی است. بنابراین باید تعداد ترکیب ها را محدود کرد. یک راه حل مناسب برای حل این مشکل در نظر گرفتن حد آستانه (α) برای کاهش تعداد ترکیب - ها و کشف کلمه های پر اهمیت می باشد. به این صورت

³⁴ Repeat of Word Polarity (RWP)

³⁵ Total Probability of Word Polarity (TPWP)

³⁶ Majority voting



مدل برچسب نهایی برای جمله مربوطه بر اساس ترکیب‌های سه‌تایی تعیین می‌گردد.

برخی از جملات دادگان آموزشی دو کلمه‌ای هستند و در قسمت استخراج قوانین سه کلمه‌ای حضور ندارند. برای تعیین برچسب مربوط به این جملات از برچسب استخراجی از قوانین دو کلمه‌ای استفاده می‌شود.

۲-۵- ارزیابی

مهم‌ترین پارامتر ارزیابی در تحقیقات مربوط به عقیده‌کاوی پارامتر صحت^{۳۷} است [۱۵]. محاسبه‌ی $FN^{۳۸}$ ، $FP^{۳۹}$ ، $TN^{۴۰}$ و $TP^{۴۱}$ ، دقت^{۴۲}، فراخوان^{۴۳} و امتیاز^{۴۴} F برای هر کلاس نیز بسیار رایج است [۱۴، ۷]. اگر TP تعداد نمونه‌های مثبتی باشد که به درستی مثبت پیش‌بینی شده‌اند، TN بیانگر تعداد نمونه‌های منفی باشد که به درستی منفی پیش‌بینی شده‌اند، FP تعداد نمونه‌های منفی باشد که به اشتباه مثبت پیش‌بینی شده‌اند و FN بیانگر تعداد نمونه‌های مثبتی باشد که به اشتباه منفی در نظر گرفته شده‌اند، در این صورت معیارهای ارزیابی صحت، دقت، فراخوان و امتیاز F به ترتیب توسط روابط ۱ تا ۴ محاسبه می‌شود:

$$صحت = \frac{TP+TN}{TP+TN+FP+FN} \quad (۱)$$

$$دقت = \frac{TP}{TP+FP} \quad (۲)$$

$$فراخوان = \frac{TP}{TP+FN} \quad (۳)$$

$$امتیاز F = \frac{2 \times دقت \times فراخوان}{دقت + فراخوان} \quad (۴)$$

در محاسبه مقادیر معیارهای دقت، فراخوان و امتیاز F در مسائل چند کلاسه از استراتژی‌های آماری

قرار می‌گیرد. در این پردازش، هر ترکیب دوتایی در کیسه کلمه‌های دوتایی جست‌وجو شده و با توجه به مقادیر محاسبه شده در مورد مثبت، منفی و یا خنثی بودن آن تصمیم‌گیری می‌شود.

در نهایت با در نظرگیری α ، برای هر جمله تعدادی دوکلمه‌ای با برچسب مشخص وجود دارد. در هر دو مدل برچسب نهایی برای جمله مربوطه بر اساس ترکیب‌های دوتایی تعیین می‌گردد.

برخی از جملات دادگان آموزشی یک کلمه‌ای هستند و در قسمت استخراج قوانین دو کلمه‌ای حضور ندارند. برای تعیین برچسب مربوط به این جملات از برچسب استخراجی از قوانین تک کلمه‌ای استفاده می‌شود.

۲،۴،۲ قانون مبتنی بر کلمات سه‌تایی

برای استخراج قوانین مبتنی بر کلمه‌های سه‌تایی از کیسه کلمه‌های سه‌تایی استفاده می‌شود. تمامی ترکیب‌های سه‌تایی از کلمه‌های موجود در جمله‌های داده آموزشی استخراج می‌شود. برای هر جمله به صورت مجزا ترکیب‌های سه‌تایی مربوطه مورد پردازش قرار می‌گیرد. در این پردازش، هر ترکیب سه‌تایی در کیسه کلمه‌های سه‌تایی جست‌وجو شده و با توجه به مقادیر محاسبه شده در مورد مثبت، منفی و یا خنثی بودن آن تصمیم‌گیری می‌شود.

در نهایت، با در نظرگیری α ، برای هر جمله تعدادی سه‌کلمه‌ای با برچسب مشخص وجود دارد. در هر دو

⁴¹ True Positive

⁴² Precision

⁴³ Recall

⁴⁴ F-score

³⁷ Accuracy

³⁸ False Negative

³⁹ False Positive

⁴⁰ True Negative



یکی از پارامترهای مهم در بررسی نتایج آزمایش مقدار تعیین شده برای α می باشد. این مقدار برای کاهش تعداد ترکیب های کلمه ها می باشد. با تغییر دادن این مقدار می توان معیارهای ارزیابی را با هم مقایسه کرد.

۱-۳- مدل RWP

نتایج مربوط به تاثیر مقادیر α در معیار ماکرو F، در جدول (۵) نشان داده شده است. با توجه به این نتایج می توان α بهینه را برای هر روش برآورد کرد. در این آزمایش از دادگان آموزشی استفاده شده است. برای راحتی نمایش روش ۱ نماینده روش مبتنی بر قوانین کلمه های تکی، روش ۲ نماینده روش مبتنی بر قوانین کلمه های دوتایی و روش ۳ نماینده روش مبتنی بر قوانین کلمه های سه تایی است.

(جدول-۵): نتایج بدست آمده از ماکرو F در مدل RWP در α

متفاوت

(Table-5): Results obtained from Macro F_score in the RWP model at different α

ماکرو F α	روش ۱	روش ۲	روش ۳	روش تلفیقی
۳	۳۹/۱۷٪	۷۴/۴۹٪	۸۶/۱۲٪	۷۵/۵۷٪
۵	۳۷/۲۷٪	۷۰/۷۲٪	۸۶/۷۸٪	۷۲/۴۹٪
۵۰	۳۲/۸۴٪	۷۸/۹۹٪	۹۳/۵۷٪	۸۰/۵۲٪
۱۰۰	۳۳/۲۱٪	۸۲/۹۹٪	۹۳/۹۸٪	۸۴/۳۵٪
۳۵۰	۴۳/۲۵٪	۸۶/۶۲٪	۹۴/۵۵٪	۸۸/۳۷٪
۵۵۰	۴۸/۸۸٪	۸۶/۸۴٪	۹۴/۶۰٪	۸۹/۱۹٪
۸۵۰	۵۵/۶۲٪	۸۶/۸۹٪	۹۴/۶۶٪	۸۹/۸۴٪
۱۰۵۰	۵۷/۴۳٪	۸۶/۹۳٪	۹۴/۶۸٪	۹۰/۰۴٪
۱۳۵۰	۵۹/۸۲٪	۸۶/۹۵٪	۹۴/۶۹٪	۹۰/۳۱٪
۱۵۵۰	۶۰/۳۲٪	۸۷/۳۹٪	۹۵/۰۲٪	۹۰/۶۸٪
۱۸۵۰	۶۴/۶۰٪	۸۷/۴۰٪	۹۵/۰۳٪	۹۱/۸۱٪
۲۰۰۰	۶۸/۷۶٪	۸۶/۹٪	۹۵/۲۹٪	۹۱/۸۱٪

با توجه به نتایج با افزایش مقدار آستانه مقادیر ماکرو F تغییر کرده است. این تغییر در همه روش ها به

میکرو و ماکرو استفاده می شود. از اینرو مشابه [۱۶] مقادیر میکرو و ماکرو به ازای هر معیار محاسبه شده است. در میانگین گیری ماکرو ابتدا بر اساس مقادیر TP، TN، FP و FN معیارهای ارزیابی به ازای هر کلاس بطور جداگانه محاسبه شده و سپس از آن مقادیر، میانگین گرفته می شود. در حالیکه در میانگین گیری میکرو، ابتدا از تمامی مقادیر TP، TN، FP و FN میانگین گرفته می شود و سپس معیارهای ارزیابی محاسبه خواهند شد. روابط ۵ و ۶ نحوه محاسبه میانگین گیری ماکرو و میکرو را نشان می دهد:

(۵)

میانگین گیری ماکرو

$$= \sum_{i=1}^k (TP_i, TN_i, FP_i, FN_i) \text{ به ازای } (TP_i, TN_i, FP_i, FN_i)$$

(۶)

میانگین گیری میکرو

$$= \left(\sum_{i=1}^k TP_i, \sum_{i=1}^k TN_i, \sum_{i=1}^k FP_i, \sum_{i=1}^k FN_i \right) \text{ به ازای } (TP_i, TN_i, FP_i, FN_i)$$

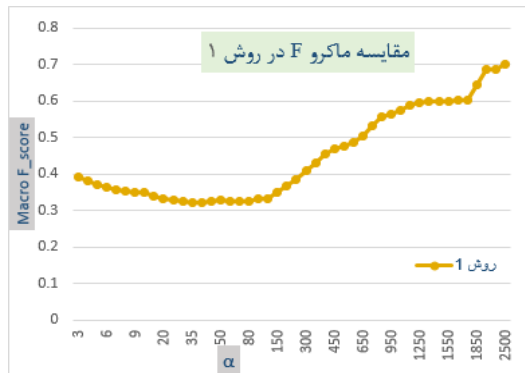
در این روابط K بیانگر تعداد کلاس ها می باشد. بدین ترتیب در این پژوهش معیارهای صحت، دقت ماکرو، فراخوان ماکرو و ماکرو F به عنوان معیار ارزیابی استفاده شده است.

۳- نتایج آزمایش

نتایج حاصل از آزمایش ها بر روی دادگان در هر دو مدل نشان می دهد که استفاده از تلفیق نتایج حاصل از سه سطح از قوانین سبب بهبود دقت شده است. این نتایج نشان می دهد که هر چه تعداد کلمه ها برای ساختن ترکیب بیشتر باشد بهتر می توان ارتباط بین کلمه ها و قطبیت جمله ها را کشف کرد. در ادامه نتایج مربوط به هر مدل شرح داده شده است.

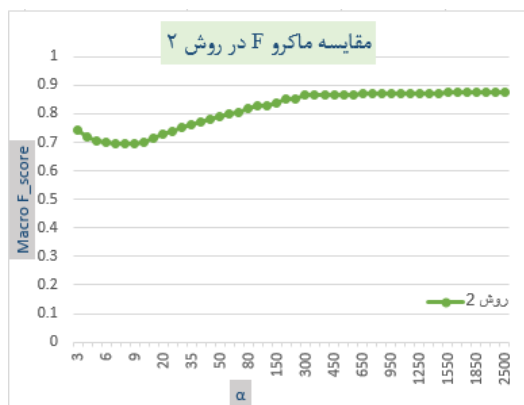


شکل‌های (۵-۲) مقایسه معیار ماکرو F در انواع روش‌های مبتنی بر قانون را در مدل RWP نشان می‌دهند.



(شکل-۲): مقایسه ماکرو F در روش مبتنی بر قوانین کلمه‌های تکی

(Figure-2): Comparison of Macro F_score in the method based on single word rules



(شکل-۳): مقایسه ماکرو F در روش مبتنی بر قوانین کلمه‌های دوتایی

(Figure-3): Comparison of Macro F_score in a dual-word rule-based method

صورت یکنواخت نمی‌باشد. در اکثریت روش‌ها این میزان در ابتدا کاهش و در ادامه افزایش یافته است.

با بهره‌گیری از α بهینه استخراج شده در آزمایش قبل می‌توان معیارهای ارزیابی را برای دادگان آزمون^{۴۵} مورد بررسی قرار داد. جدول (۶) نتایج بدست آمده از پیاده‌سازی روش پیشنهادی در مدل RWP با استفاده از α بهینه بروی دادگان آزمون را نشان می‌دهد.

(جدول-۶): نتایج بدست آمده از پیاده‌سازی روش پیشنهادی در مدل RWP با بهره‌گیری از α بهینه

(Table-6): The results obtained from the implementation of the proposed method in the RWP model using optimal α

$\alpha = 2000$				
	صحت	دقت ماکرو	فراخوان ماکرو	ماکرو F
روش مبتنی بر قوانین کلمه-های تکی	۸۲/۱۴٪	۸۵/۰۴٪	۷۷/۳۹٪	۸۰٪
روش مبتنی بر قوانین کلمه-های دوتایی	۹۳/۲۸٪	۹۴/۶۶٪	۹۱/۰۲٪	۹۲/۵۸٪
روش مبتنی بر قوانین کلمه-های سه‌تایی	۹۷/۹۷٪	۹/۵۶٪	۹/۹۲٪	۹/۶۸٪
روش تلفیقی	۹۶/۰۵٪	۹۶/۸۰٪	۹۴/۳۶٪	۹۵/۴۴٪

با بررسی نتایج موجود در جدول مشخص است که روش مبتنی بر قوانین حاصل از ترکیب سه‌تایی کلمه‌ها کارایی بهتری نسبت به دو روش مبتنی بر کلمه‌های تکی و دوتایی داشته است. روش تلفیقی نیز به نسبت کارایی مناسبی در زمینه عقیده‌کاو به زبان فارسی دارد.

⁴⁵ Test data



در بازه انتهایی میزان این معیار تقریباً به یک مقدار ثابت رسیده است.

با توجه به شکل (۲)، در بازه ابتدایی یعنی α بین ۳ تا ۲۵۰، میزان معیار ارزیابی کاهش یافته است. ولی در ادامه این میزان افزایش داشته است.

با توجه به شکل (۵)، در مقادیر ابتدایی α افت ناچیزی در ماکرو F وجود دارد ولی در ادامه روند، با افزایش α میزان ماکرو F نیز افزایش یافته است.

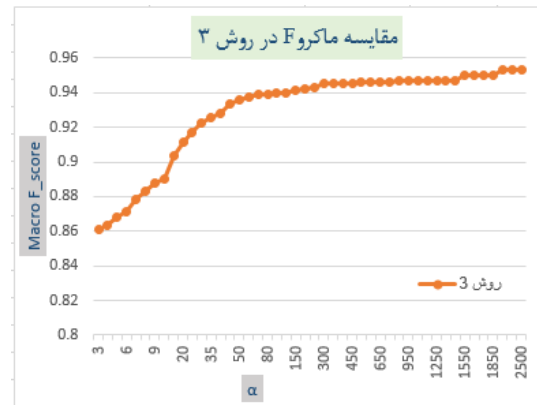
۳-۲- مدل TPWP

در مدل TPWP نیز تغییر مقادیر α در معیارهای ارزیابی تاثیر گذار بوده است. نتایج موجود در جدول (۷)، تاثیر این مقادیر را در معیار ماکرو F به خوبی نشان می دهد. در این مدل نیز روش ۳ دارای عملکرد بهتری بوده است.

(جدول-۷): نتایج بدست آمده از ماکرو F در مدل TPWP در α متفاوت

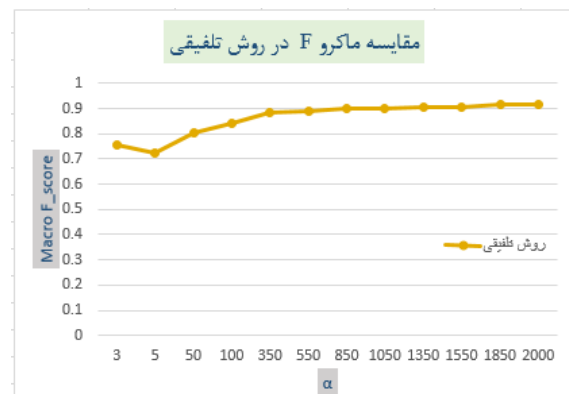
(Table-7): Results obtained from Macro F_score in the TPWP model at different α

ماکرو F α	روش ۱	روش ۲	روش ۳	روش تلفیقی
۳	۴۵/۲۵٪	۷۷/۲۴٪	۸۷/۵۰٪	۷۸/۸۷٪
۵	۴۲/۵۹٪	۷۳/۱۲٪	۸۷/۷۴٪	۷۵/۶۵٪
۵۰	۳۵/۷۰٪	۷۹/۶۶٪	۹۴/۰۸٪	۸۱/۳۹٪
۱۰۰	۳۴/۷۸٪	۸۳/۴۱٪	۹۴/۳۸٪	۸۴/۸۰٪
۳۵۰	۴۴/۱۵٪	۸۷/۰۸٪	۹۴/۹۷٪	۸۸/۸۳٪
۵۵۰	۴۹/۶۳٪	۸۷/۲۸٪	۹۴/۹۹٪	۸۹/۶۰٪
۸۵۰	۵۵/۹۶٪	۸۷/۲۸٪	۹۵/۰۱٪	۹۰/۱۸٪
۱۰۵۰	۵۷/۷۴٪	۸۷/۳۰٪	۹۵/۰۱٪	۹۰/۳۷٪
۱۳۵۰	۵۹/۸۲٪	۸۷/۳۳٪	۹۵/۰۲٪	۹۰/۶۲٪
۱۵۵۰	۵۹/۹۶٪	۸۷/۳۹٪	۹۵/۰۲٪	۹۰/۶۸٪
۱۸۵۰	۶۴/۶۰٪	۸۷/۴۰٪	۹۵/۰۳٪	۹۱/۲۳٪
۲۰۰۰	۶۸/۷۶٪	۸/۶۹٪	۹۵/۲۹٪	۹۱/۸۱٪



(شکل-۴): مقایسه ماکرو F در روش مبتنی بر قوانین کلمه های سه تایی

(Figure-4): Comparison of Macro F_score in a triple-word rule-based method

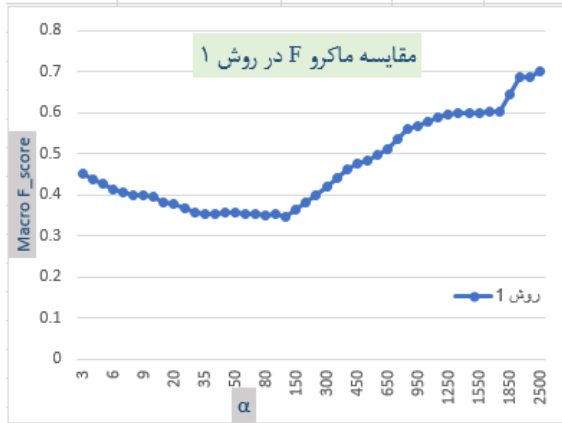


(شکل-۵): مقایسه ماکرو F در روش تلفیقی

(Figure-5): Comparison of Macro F_score in the compilation method

با توجه به شکل (۴)، نمودار حاصل از روش ۳ نسبت به سایر نمودارها با تغییر مقدار α دارای افت مقدار ماکرو F نمی باشد. از این نمودار استنتاج می شود که مقدار α تاثیر کمتری نسبت به دیگر روش ها بر روی میزان معیار مورد استفاده داشته است.

با توجه به شکل (۳)، در بازه ابتدایی یعنی α بین ۳ تا ۲۰، میزان دقت کاهش یافته است. اما در ادامه روند افزایش مقدار α ، میزان ماکرو F نیز افزایش یافته است.

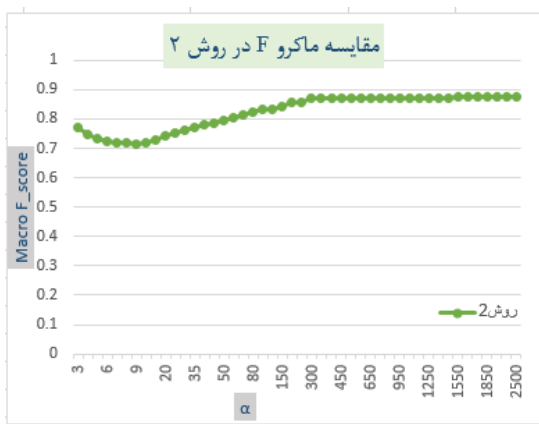


(شکل-۶): مقایسه ماکرو F در روش مبتنی بر قوانین کلمه های تکی

(Figure-6): Comparison of Macro F_score in the method based on single word rules

در این مدل تقریباً همه شکل‌ها تحلیل یکسانی دارند. در یک بازه مشخص کاهش دقت و در ادامه افزایش دقت مشاهده می‌شود.

با توجه به شکل (۶)، در بازه بین ۳ تا ۲۵۰ میزان ماکرو F کاهش یافته است. در بازه ۲۵۰ به بعد میزان آن افزایش یافته است.



(شکل-۷): مقایسه ماکرو F در روش مبتنی بر قوانین کلمه های دوتایی

(Figure-7): Comparison of Macro F_score in a dual-word rule-based method

جدول (۸) نتایج بدست آمده از پیاده‌سازی روش پیشنهادی در مدل TPWP با استفاده از α بهینه بروی دادگان آزمون را نشان می‌دهد.

(جدول-۸): نتایج بدست آمده از پیاده‌سازی روش پیشنهادی در

مدل TPWP با بهره‌گیری از α بهینه

(Table-8): The results obtained from the implementation of the proposed method in the TPWP model using optimal α

$\alpha = 2000$				
	صحت	دقت ماکرو	فراخوان ماکرو	ماکرو F
روش مبتنی بر قوانین کلمه-های تکی	۸۲/۱۴٪	۸۵/۰۴٪	۷۷/۳۹٪	۸۰٪
روش مبتنی بر قوانین کلمه-های دوتایی	۹۳/۲۸٪	۹۴/۶۶٪	۹۱/۰۲٪	۹۲/۵۸٪
روش مبتنی بر قوانین کلمه-های سه تایی	۹۸/۲۲٪	۹۸/۸۳٪	۹۶/۹۱٪	۹۷/۷۹٪
روش تلفیقی	۹۷/۴۰٪	۹۷/۹۸٪	۹۵/۶۶٪	۹۶/۷۰٪

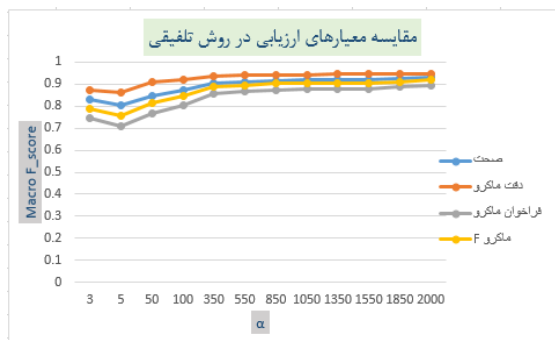
با مقایسه نتایج جدول (۶) و جدول (۸) مشخص است که هر دو مدل کارایی تقریباً مشابهی داشته‌اند. در این مدل نیز روش مبتنی بر قوانین حاصل از کلمه-های سه تایی کارایی بهتری نسبت به دو روش دیگر داشته است. روش تلفیقی از هر سه روش نیز مانند مدل قبل به نتیجه خوبی رسیده است.

شکل‌های (۶-۹)، مقایسه معیار ماکرو F در انواع روش‌های مبتنی بر قانون را در مدل TPWP نشان داده است.



با توجه به شکل (۱۰)، به صورت کلی در بازه ابتدایی از مقادیر α ، روش تلفیقی در مدل TPWP دارای کارایی بهتری بوده است.

در شکل (۱۱)، معیارهای ارزیابی در روش تلفیقی در مدل TPWP با هم مقایسه شده‌اند.



(شکل-۱۱): مقایسه معیارهای ارزیابی در روش تلفیقی در مدل TPWP

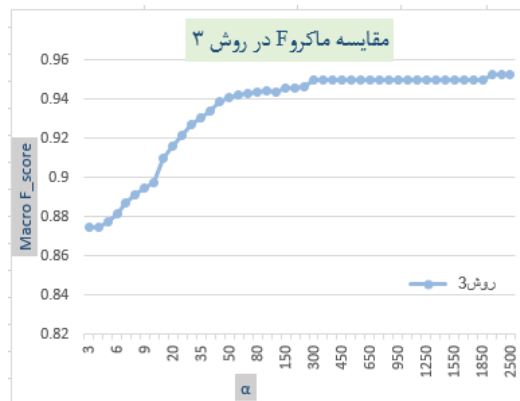
(Figure-11): Comparison of evaluation criteria in the combined method in the TPWP model

با مشاهده شکل (۱۱) مشخص است که معیار دقت نسبت به سایر معیارها دارای مقادیر بیشتری است. معیار بازیابی نیز دارای کمترین مقادیر می‌باشد.

۴- نتیجه گیری

یکی از چالش‌های مطرح عقیده‌کاوی در زبان فارسی عدم وجود دادگان مناسب در این حوزه می‌باشد. در این پژوهش دادگانی با استفاده از گروه‌های تلگرام فارسی ساخته شده است. با رفع این چالش می‌توان از این دادگان در دیگر پژوهش‌ها نیز بهره برد.

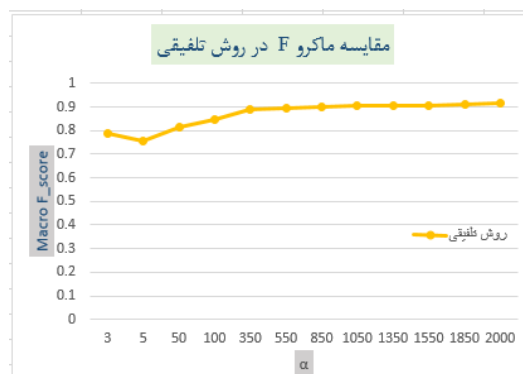
نتایج حاصل از استخراج قوانین نشان می‌دهد که روش مبتنی بر قوانین کلمه‌های سه‌تایی دارای عملکرد بهتری بوده است. علاوه بر این، نتایج حاصل از تلفیق سه سطح از قوانین استخراج شده نشان می‌دهد که دقت در روش تلفیقی دارای بهبود بوده است. یکی از



(شکل-۸): مقایسه ماکرو F در روش مبتنی بر قوانین کلمه‌های

سه‌تایی

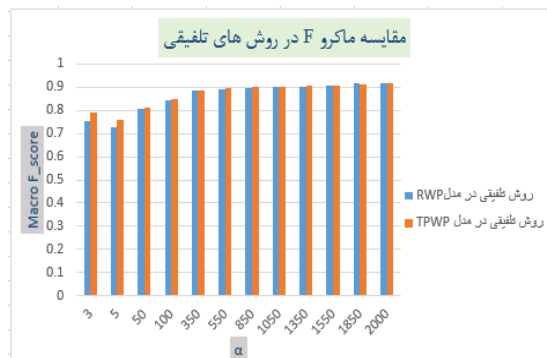
(Figure-8): Comparison of Macro F_score in a triple-word rule-based method



(شکل-۹): مقایسه ماکرو F در روش تلفیقی

(Figure-9): Comparison of Macro F_score in the compilation method

شکل (۱۰) مقایسه بین روش‌های تلفیقی در دو مدل را نشان می‌دهد.



(شکل-۱۰): مقایسه روش تلفیقی در دو مدل

(Figure-10): Comparison of the combined method in two models



Intelligent Systems and Applications 4, no. 10, pp. 1, 2012.

- [11] Basiri, Mohammad Ehsan, Ahmad Reza Naghsh-Nilchi, and Nasser Ghassem-Aghaee. "A framework for sentiment analysis in persian." OPEN TRANSACTIONS ON INFORMATION PROCESSING 1, PP. 14, 2014.
- [۱۲] پ. ف. مجازی و ب. خ. جوان. (۱۳۹۶) ۱۲/۷، تلگرام چقدر ایرانیها را به خود مشغول کرده است؟
Available: <http://www.yjc.ir/fa/news/5921627>.
- [13] <http://ml.yazd.ac.ir>
- [14] J. Park, V. Barash, C. Fink, and M. Cha. Emoticon style: Interpreting differences in emoticons across cultures. In International AAAI Conference on Weblogs and Social Media (ICWSM), 2013.
- [15] Kharde, Vishal, and Prof Sonawane. "Sentiment analysis of twitter data: A survey of techniques." arXiv preprint arXiv: 1601.06971, 2016.
- [16] Vincent Van Asch, "Macro- and micro-averaged evaluation measures". Semantic scholar, 2013.



مینا کشاورزی فراشاه تحصیلات

خود را در مقطع کارشناسی در رشته مهندسی کامپیوتر گرایش نرم افزار در سال ۱۳۹۴ در دانشگاه یزد به پایان رساند. وی توانست با استفاده از سهمیه استعداد درخشان مقطع

کارشناسی ارشد خود را در همین رشته در دانشگاه یزد ادامه دهد و هم اینک مشغول تحصیل می باشد. زمینه های تحقیقاتی مورد علاقه نامبرده عبارتند از: یادگیری ماشین، عقیده کاوی در شبکه های اجتماعی، متن کاوی و داده کاوی.

نشانی رایانامه ایشان عبارت است از:

m.keshavarzi@stu.yazd.ac.ir



دکتر محمدعلی زارع چاهوکی

مدارک کارشناسی ارشد و دکترای خود را در رشته مهندسی کامپیوتر نرم افزار از دانشگاه تربیت مدرس به ترتیب در سال های ۱۳۸۳ و ۱۳۹۱ و نیز مدرک کارشناسی خود را در

رشته مهندسی کامپیوتر نرم افزار از دانشگاه شهید بهشتی در

پارامترهای ضروری در پیاده سازی روش پیشنهادی α است. انتخاب α نقش مهمی در میزان معیارهای ارزیابی داشته است.

نتایج بدست آمده از این پژوهش می تواند با روش های مبتنی بر یادگیری ماشین ترکیب شود. استفاده از روش پیشنهادی در کنار روش های یادگیری ماشین می تواند بهبود قابل توجهی در معیارهای ارزیابی داشته باشد

۵- مراجع

- [1] Asghar MZ, Khan A, Ahmad S, Qasim M, Khan IA. Lexicon-enhanced sentiment analysis framework using rule-based classification scheme. PLoS ONE 12(2): e0171649. Doi: 10.1371/journal.pone.0171649, 2017.
- [2] Khan, Farhan Hassan, Saba Bashir, and Usman Qamar. "TOM: Twitter opinion mining framework using hybrid classification scheme." Decision Support Systems 57, pp. 245-257, 2014.
- [3] Lee, Dongjoo, Ok-Ran Jeong, and Sang-goo Lee. "Opinion mining of customer feedback data on the web." Proceedings of the 2nd international conference on Ubiquitous information management and communication. ACM, 2008.
- [4] K. Khan, B. B. Baharudin, A. Khan, and Fazal-e-Malik, "Mining Opinion from Text Documents: A Survey " 3rd IEEE International Conference on Digital Ecosystems and Technologies, 2009.
- [5] M.A.AZH Khan, "Combining lexicon-based and learning-based methods for Twitter sentiment analysis," International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCE), pp. 89-91, 2015.
- [6] Saif, Hassan, et al. "Contextual semantics for sentiment analysis of Twitter." Information Processing & Management. vol.52, no.1, pp.5-19, 2016.
- [7] Giachanou, Anastasia, and Fabio Crestani. "Like it or not: A survey of twitter sentiment analysis methods." ACM Computing Surveys (CSUR), vol. 49, no. 2, pp.28, 2016.
- [8] B. Liu, Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, 2012.
- [9] Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey." Ain Shams Engineering Journal, vol. 5, no. 4, pp. 1093-1113, 2014.
- [10] Kumar, Akshi, and Mary Sebastian Teeja. "Sentiment analysis: A perspective on its past, present and future." International Journal of



سال ۱۳۷۸ دریافت نمودند. ایشان هم اکنون استادیار و هیأت علمی گروه مهندسی کامپیوتر دانشگاه یزد هستند که زمینه های تحقیقاتی مورد علاقه ایشان شامل بینایی ماشین، مهندسی نرم افزار، یادگیری ماشین، عقیده کاوی از داده های شبکه های اجتماعی و یادگیری عمیق می باشد.

نشانی رایانامه ایشان عبارت است از:

chahooki@yazd.ac.ir

نشانی وب سایت ایشان عبارت است از:

<https://pws.yazd.ac.ir/chahooki>