

آنالیز احساسی متون فارسی

شکوفه دشتبانی مشکانی^۱، عبدالحمید پیله ور^۲

آزمایشگاه مهندسی زبان، گروه کامپیوتر، دانشکده مهندسی، دانشگاه بوعلی سینا همدان

^۱S.Dashtbani@basu.ac.ir , ^۲pilevar@basu.ac.ir

چکیده:

هدف از عرضه این مقاله بررسی راهکار جدیدی به منظور آنالیز متون فارسی می باشد. در این مقاله آنالیز احساسی متون فارسی مورد بررسی قرار می گیرد و خصایص و ویژگی های روش پیشنهاد شده در این مقاله با روش های موجود در زبان انگلیسی مقایسه می گردد. اخیراً آنالیز احساسی متون در زبان های دیگر به ویژه زبان انگلیسی مورد مطالعه قرار گرفته است و روش هایی نیز به منظور آنالیز سریع متون انگلیسی ارائه شده است. روشی که در این مقاله مورد استفاده قرار گرفته است، تا حدودی نتیجه مشابه با روش های مورد استفاده در زبان انگلیسی داشته است. در این روش آنالیز احساسی با ساخت دسته های احساسی صورت می گیرد که روش ساخت دسته های احساسی بر اساس شباهت معنایی می باشد. استفاده از شباهت معنایی برای ساخت دسته های احساسی در زبان فارسی و آنالیز احساسی متون فارسی به طور کلی در حدود ۸۰ درصد موارد موفقیت آمیز بوده است.

کلمات کلیدی

هوش مصنوعی، شباهت معنایی^۱، شباهت کلمات^۲، دسته های احساسی^۳

۱- مقدمه:

پردازش زبان طبیعی یکی از شاخه های هوش مصنوعی می باشد که پردازش زبان فارسی در این مقاله مورد تحقیق قرار گرفته است. آنالیز احساسی ساختار متن از آنجا حائز اهمیت است که می تواند ابهامات را تا حدودی کاهش دهد. تحقیقات گسترده ای در زمینه آنالیز احساسی متون انجام شده است، استفاده از آنالیز احساسی به منظور استخراج اندیشه^۴ [1] و همچنین طبقه بندی و آنالیز احساسی سخنان گفته شده^۵ [2] از قبیل کارهای انجام شده در این زمینه می باشد.

یکی از تکنیک های اساسی که در این مقاله مورد استفاده قرار گرفته است، استفاده از شباهت کلمات در زبان فارسی می باشد که از آن برای پیدا کردن شباهت لغات متن فارسی با دسته های احساسی مورد استفاده قرار گرفته است. راهکارهایی تاکنون برای زبان انگلیسی ارائه گردیده است، از قبیل بدست آوردن شباهت با استفاده از بردار ویژگی^۶ [3]، تخمین بر اساس فاصله بین دو نود^۷ [4]، بر اساس ارتباط مفهومی^۸ [5]، استفاده از منطق فازی [6] و همچنین شباهت معنایی را می توان بر اساس محاسبات

آماري^۹ [7] نیز محاسبه کرد. ساختن دسته های احساسی نیازمند مطالعه ی دقیق در دستور زبان فارسی و یا استفاده از یک متخصص در زمینه ادبیات فارسی می باشد که در این مقاله 40 دسته احساسی در زبان فارسی با استفاده از کتب دستور زبان فارسی [8] معرفی می شود که از این 40 دسته احساسی پس از تجزیه متن ورودی به کلمات تشکیل دهنده اش، برای آنالیز متن استفاده می گردد.

ابهامات متن با توجه به اینکه لغت ورودی در چه دسته های احساسی می تواند قرار بگیرد، مدیریت خواهند شد. رفع ابهام از این جهت است که علاوه بر بیان احساس، کلمه در زبان طبیعی را نیز بیان می کند [9]. علاوه بر طبقه بندی معنایی، در اینجا از چگالی عددی کلمات در متن ورودی نیز برای مدیریت ابهامات استفاده می شود. تمامی کلمات احساسی متن به همراه چگالی عددی و عددی که نشان دهنده میزان شباهت کلمه ورودی به دسته های احساسی می باشد را در یک فایل پیوست به متن ورودی قرار می دهیم، در نهایت با توجه به آن متن ورودی را مورد آنالیز قرار می دهیم. در زبان انگلیسی اخیراً یکسری محاسبات احساسی نیز ارائه گردیده است [10]. در آخر هم پیشنهاد هایی در