

چکیده

آگاهی از نظرات و دیدگاه سایر افراد نقش اساسی در فرایند تصمیم‌گیری مشتریان ساده تا مدیران سطح بالای شرکت‌های تولیدکننده و سازمان‌های مختلف دارد. امروزه با پیدایش وب ۲,۰ و گسترش شبکه‌های اجتماعی، حجم زیادی از متون مربوط به نظرات افراد ایجاد شده است. ولی کاوش در حجم انبوه مستندات، منابع نظرات متعدد و وجود نظرهای مغایر درباره یک موجودیت، فرایند استخراج و پردازش نظرات را بسیار دشوار ساخته است. بنابراین نیاز به روشهایی برای کاوش و خلاصه‌سازی نظرات موجود احساس می‌شود. بدین منظور در دهه اخیر گرایش جدیدی در علم پردازش زبان‌های طبیعی به نام نظر کاوی ایجاد شده است. کارهای پژوهشی اولیه در این حوزه، تنها محدود به استخراج و تحلیل حس مثبت یا منفی افراد، از روی اطلاعات متنی بودند. در سالهای اخیر، تحقیقات جدید نظر کاوی بیشتر به سمت نظر کاوی مبتنی بر ویژگی (جنبه) پیش رفته است. هدف اصلی ما در این پژوهش، نظر کاوی در سطح ویژگی و خلاصه‌سازی ساخت‌یافته نظرات (میزان حس کاربران به تفکیک ویژگی‌های متون) است. روشهای موجود برای تشخیص حس نظرات، در برخورد با پیچیدگی‌های زبانی و مفهومی، ضعف‌های زیادی دارند.

پس با توجه به پیچیدگی‌های خلاصه‌ساز ساخت‌یافته نظرات، در این پژوهش از رویکرد معنایی برای استخراج ویژگی‌ها، تشخیص حس مرتبط با هر ویژگی و نهایتاً بیان روابط و نمایش نتایج استفاده شده است. ایده اصلی این پژوهش در نظر گرفتن معنا (دانش) برای تشخیص حس در متن با استفاده از یک چارچوب معنایی جهت استفاده در تمام مراحل خلاصه‌سازی نظرات بصورت یکپارچه است. بدین منظور، روشی نیمه خودکار برای ایجاد پایگاه شناخت نظرات پیشنهاد می‌شود. هدف از ارائه این چارچوب، تبدیل حجم انبوه متن نظرات از قالب غیرساخت‌یافته متنی به قالب داده‌های معنایی ساخت‌یافته در زمان قابل قبول است. استفاده از پایگاه شناخت نظرات پیشنهادی، کمک می‌کند تا ویژگی‌های مختلف و ارتباط بین آنها با موجودیت اصلی و همچنین حس بیان شده برای هر ویژگی به خوبی مشخص گردند.

قبل از تولید پایگاه شناخت نظرات، یک مدل مفهومی نظرات (شمای پایگاه شناخت) مستقل از زبان و دامنه نظرات، جهت نگهداری نظرات در قالب ساخت‌یافته، پیشنهاد می‌شود. در مرحله بعد، زبان مقصد انتخاب شده و یک روش نیمه خودکار (با استفاده از دانش اولیه) برای تشکیل پایگاه شناخت نظرات از روی مجموعه مستندات (در یک حوزه خاص)، پیشنهاد شده است. سپس از پایگاه شناخت نظرات تولید شده، برای کشف ویژگی‌ها، تحلیل حس هر ویژگی و گروه‌بندی آنها استفاده می‌شود. نهایتاً خلاصه‌ی نظرات به شکل ساخت‌یافته در قالب داده‌های معنایی بیان شده‌اند. نتایج ارزیابی‌ها نشان داد که چارچوب ارائه شده کیفیت مناسبی برای خلاصه‌سازی ساخت‌یافته نظرات دارد.

همچنین در راستای هدف اصلی رساله، پیکره نظرات فارسی، مجموعه واژگان حسی عمومی و خاص (برای کالاهای تجاری)، مجموعه‌ای از ابزارهای مختلف پردازش متن فارسی مورد نیاز برای نظر کاوی و همچنین شبکه واژگان فردوس‌نت برای زبان فارسی تهیه شده است.