

دسته‌بندی یا دیگر کاربردهای متن کاوی و پردازش زبان طبیعی دارد. با توجه به اهمیت مجموعه واژگان حسی و فاز پیش‌پردازش در تحلیل نظرات [۱۸۳]؛ لازم است تا مجموعه‌ای از ابزارهای کاربردی پردازش متون زبان فارسی برای استفاده در قسمت‌های مختلف نظرکاوی تهیه شود.

متاسفانه ابزارهای استاندارد پیش‌پردازش ایجاد شده برای متون زبان فارسی از قبیل [۲۰۴-۲۰۶] بصورت رایگان منتشر نشده‌اند. برخی از ابزارهای کد باز<sup>۱</sup> پیش‌پردازش موجود از قبیل [۲۰۷-۲۰۹] نیز از دقت مناسبی برخوردار نیستند. بطور کلی کمبود منابع استاندارد و با کیفیت (از قبیل مجموعه واژگان حاوی حس استاندارد) و برخی از ابزارهای پردازش زبان (از قبیل NER، تشخیص مرجع ضمائر، پارسر عمقی و برجسب‌گذار معنایی نقش‌ها) و سایر پیچیدگی‌های نحوی و گرامری زبان فارسی، بخصوص در متن نظرات که به شکل محاوره‌ای و عامیانه نوشته می‌شود، مانع بزرگی برای انجام مطالعات و پیاده‌سازی خلاصه‌ساز نظرات برای زبان فارسی است. البته در این رساله بعضی از ابزارهای پردازش متن و منابع مورد نیاز این زمینه تحقیقاتی، تهیه و منتشر شده است.

### ۳-۵- کارهای آتی

گرچه نتایج ارزیابی رویکرد پیشنهادی رساله رضایت‌بخش و قابل قبول است، ولی برای رفع چالش‌های اشاره شده در این پژوهش، راه‌حلهایی برای توسعه و بهبود رویکرد ارائه شده وجود دارد. در این بخش، مهمترین گام‌هایی که در ادامه مسیر این پژوهش قابل انجام است، در ذیل ارائه خواهد شد.

- تحلیل نظرات مقایسه‌ای: نظرات مقایسه‌ای حاوی اطلاعات بسیار مفیدی از تجربیات کاربران در دو یا چند زمینه هستند و تحلیل این نوع نظرات می‌تواند باعث افزایش کیفیت تحلیل نظرات بخصوص برای خلاصه‌سازی نظرات شود. البته در شمای پایگاه شناخت نظرات پیشنهاد شده امکان بیان و انتشار این نوع نظرات وجود دارد.

---

<sup>۱</sup> Open Source

- تحلیل نظرات غیرمستقیم: روش‌های تحلیل و تشخیص حس برای نظرات غیرمستقیم (کنایه‌ای) و مشروط (بیان حس با قید شرطی یا اشاره به زمان آینده) نیازمند به تحلیل‌های عمقی زبان‌شناسی یا وجود پیکره بزرگ برای شناسایی و تحلیل هستند.
- حذف نظرات هرز و نامرتبط: تشخیص و حذف نظرات هرز (اسپم)، بی‌معنی، پرسش‌ها یا نظرخواهی کاربران از یکدیگر در بخش نظرات، نظرات نوشته شده در پاسخ سایر نظرات یا بحث با سایر کاربران در بخش نظرات، می‌تواند تا حد قابل ملاحظه‌ای موجب افزایش دقت تحلیل حسی نظرات گردد.
- استفاده از اطلاعات پروفایل و سوابق کاربران برای تلفیق حسی و خلاصه‌سازی نظرات: با توجه به سطح خبرگی متفاوت کاربران در حوزه (دامنه) مختلف، نظرات آنها از میزان ارزشمندی یکسانی برخوردار نیستند. لذا مدل‌سازی تلفیق احساسات بیان شده در نظرات مختلف با استفاده از اطلاعات پروفایل و سوابق کاربران مخصوصاً برای خلاصه‌سازی نظرات بسیار مفید خواهد بود.
- توسعه ابزارهای پردازش متن مورد نیاز برای تحلیل عمیق متن نظرات فارسی: در این پژوهش فرض نمودیم که کاربر بدون ارجاع مستقیم به یک ویژگی، درباره ویژگی‌های چند جمله قبل، حسی را بیان نمی‌کند. علت اصلی این فرض، نبود ابزار تشخیص‌دهنده ارجاع ضمائر مناسب در زبان فارسی است. تشخیص ارجاع ضمائر می‌تواند به تعیین ویژگی‌ها و ارتباط آن با حس بیان شده در جمله فعلی کمک قابل توجهی کند.
- همچنین اصلاح ساختاری جملات محاوره‌ای با هدف استفاده از پارسر وابستگی اجزای جمله و تعیین ارتباط بین ویژگی‌ها و کلمات حسی یا تولید پیکره وابستگی متون محاوره‌ای برای آموزش پارسر وابستگی جملات محاوره‌ای می‌تواند نقش موثری در تعیین روابط بین ویژگی و عبارات حسی داشته باشد.
- بهبود روش استخراج ویژگی در رویکرد پیشنهادی (در مرحله توسعه پایگاه شناخت): همانطور که بیان شد، برای توسعه پایگاه شناخت ویژگی‌های نظرات در یک رویکرد تکراری-افزایشی، شناسایی و گروه‌بندی می‌شوند. لذا پیشنهاد می‌شود که روش یادگیری شبه ناظر ایجاد گردد که با در نظر

گرفتن خروجی‌های معتبر تکرار اول به عنوان ویژگی‌های اولیه (دانش ضمنی) بتواند برای استخراج ویژگی‌های جدید در تکرار بعدی بهره گیرد.

- استفاده از روش‌های یادگیری عمیق جهت شناسایی توأم ویژگی و حس : در سال‌های اخیر، توجه عمده‌ای به استفاده از روش‌های یادگیری عمیق برای استخراج حس هر ویژگی در جملات کوتاه موجود در شبکه‌های اجتماعی (بوئژه توئیتز) شده است. لذا پیشنهاد می‌شود، تحقیقات و تمرکز ویژه‌ای بر روی دسته از روش‌ها انجام شود.