

## عنوان مقاله:

بهبود الگوریتم خوشه بندی جهت تحلیل احساس در متون فارسی

## محل انتشار:

پنجمین کنفرانس بین المللی مهندسی برق و کامپیوتر با تاکید بر دانش بومی (سال: 1396)

تعداد صفحات اصل مقاله: 12

## نویسندگان:

شیما هاشم زاده - گروه کامپیوتر، واحد بویین زهره، دانشگاه آزاد اسلامی، بویین زهره، ایران

مهدی ملامطلبی - گروه کامپیوتر، واحد بویین زهره، دانشگاه آزاد اسلامی، بویین زهره، ایران

حسن نادری - گروه مهندسی کامپیوتر، دانشگاه علم و صنعت، تهران، ایران

## خلاصه مقاله:

امروزه با رشد سریع اینترنت و در دسترس بودن برای همه مردم دنیا، حجیم عظیمی از اطلاعات فراهم آمده است. اغلب این داده ها به صورت متنی هستند. تحلیل احساس شامل استخراج عقاید، احساسات، ارزیابی های کاربران از داده های متنی میباشد. تاکنون روشهایی به منظور دسته بندی متون به دو گروه داده های متنی حاوی نظرات مثبت و گروه حاوی نظرات منفی ارایه شده است. اغلب این روشها از یادگیری ماشین مبتنی بر ناظر بهره میبرند درحالیکه در بسیاری از موارد، ممکن است داده های لازم برای آموزش ماشین در دسترس نباشند. در این پژوهش، به تحلیل احساس درون متون پرداخته ایم. روش ارایه شده در این پژوهش از الگوریتم های خوشه بندی که در واقع همان روشهای یادگیری ماشین بدون ناظر میباشد، استفاده شده است. در این پژوهش هر متن را با استفاده از خصیصه های word-1 و word-2 مدل کرده ایم. با توجه به اینکه داده های متنی شامل حجم زیادی از خصیصه ها هستند، به منظور کاهش تعداد خصیصهها ابتدا ریشه یابی انجام شده است، سپس کلمات متوقف کننده حذف شده اند و در ادامه خصیصههایی که فرکانس وقوع آنها از یک آستانه کمتر است را حذف کرده ایم. با بکارگیری الگوریتم LDA، عناوین مهم که در دو نوع متون مثبت و منفی مجموعه خصیصهها که دارای اهمیت بیشتری هستند را استخراج کرده ایم و سپس با الگوریتم انتخاب خصیصه IGR فیلتر نهایی را روی خصیصه ها اعمال کرده ایم. از چندین مرحله فیلتر خصیصه ها استفاده شده است که خصیصههای غیر سودمند خود چندین دسته خصیصههای غیر مرتبط، خصیصههای افزونه و خصیصههای نامتعادل میباشد. در نظر گرفتن چندین مرحله فیلترسازی خصیصهها روش مناسبی به منظور پالایش انواع مختلف خصیصه های غیر سودمند میباشد. در پایان با استفاده از الگوریتم EM خوشه بندی داده ها بر روی خوشه های مثبت و منفی انجام داده ایم. نتایج نشان میدهد روش پیشنهادی در این پژوهش نسبت به روشهای پیشین بهبود قابل توجهی ارایه داده است.

## کلمات کلیدی:

متن کاوی، تحلیل احساس، انتخاب خصیصه، داده کاوی، خوشه بندی

## لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/725112>

