

## عنوان مقاله:

کاوش نظرات مقایسه ای در نقدهای محصول فارسی برای هوش رقابتی

## محل انتشار:

کنگره ملی سالانه ایده های نوین پژوهشی در علوم مهندسی و تکنولوژی، برق و کامپیوتر (سال: 1397)

تعداد صفحات اصل مقاله: 15

## نویسنده:

نوا سروش حدادی - کارشناسی ارشد مهندسی فناوری اطلاعات تجارت الکترونیک، دانشگاه آزاد واحد الکترونیکی، تهران

## خلاصه مقاله:

در میان انبوه تحقیقاتی که در دهه ی اخیر در حوزه نظرکاوی منتشر گشته اند، تعداد کمی از کارها به زبان های غیر انگلیسیمربوط میشود. در این میان، بنا به دانش نویسنده، تاکنون تنها تعداد اندکی تحقیق در زمینه ی نظرکاوی مستقیم در زبانفارسی انجام گرفته اند که آنها نیز جملات مقایسه ای را در نظر نگرفته اند. هدف این تحقیق ارایه ی یک روش دسته بندیجملات مقایسه ای و غیر مقایسه ای با در نظر گرفتن برچسب های (قیاس نامساوی، قیاس مساوی، قیاس تفضیل، غیر قیاسی و غیر مقایسه ای) نظرات سه دسته محصول در سایت فروشگاه الکترونیکی دیجیکالا، با دقت مطلوب میباشد که بتواند به عنوانفاز اول فیلترینگ، برای نظرکاوی جملات مقایسه ای در بسیاری از کارهای پردازش زبان طبیعی و هوش رقابتی، از جمله پیداکردن بینش بازار و عملکرد فروش یک شرکت در مقایسه با رقبای آن، مورد استفاده قرار بگیرد. بدین منظور، از روشهایبایدگیری ماشین استفاده کرده ایم که از موثرترین و محبوب ترین روشها در طبقه بندی متن میباشدند. در این تحقیق، دسته-بندهای k نزدیکترین همسایه، درخت تصمیم و نایویز در مرحله ی یادگیری مورد استفاده قرار گرفته اند. فرآیند روشپیشنهادی بدین صورت است که پس از پیش پردازش متن رودی، با توجه به این که مجموعه داده ی مورد استفاده در اینتحقیق خاصیت نامتوازن دارد، روش پیشنهادی با انتخاب وسیله ی یک تکنیک دسته بندی با استفاده از نمونه گیری مجدد ازداده ها (Resample) تلاش می کند که برای داده های نامتقارن مفید واقع شود و تاثیرات منفی محدودیتهای ایجاد شده را بهوسیله ی داده های نامتقارن را از بین ببرد. این کار به وسیله ی یک تکنیک دسته بندی با استفاده از نمونه گیری مجدد از داده هانجام می شود. در مرحله ی بعد با توجه به اینکه عملکرد این دسته بندهای بانظارت اغلب تحت تاثیر بردار ویژگی ورودیمتفاوت میباشدند، در این پژوهش، به عنوان جایگزینی برای بردار کلمات، مجموعه ویژگی های n-gram؛ (1-3 گرم) به عنوانبردار کلمات استخراج میشوند. سپس به منظور کاهش تعداد ویژگیها و انتخاب ویژگیهای اصلی متن، از روش انتخاب ویژگیمعیار افزایش اطلاعات، مجموعه ویژگیهای مناسب برای طبقه بندی انتخاب میشوند. در مرحله ی بعد، برای انعکاس ساختاردقیق تری از متون در بردار ویژگی، به جای موجودیت های مورد مقایسه، آنها را با عبارتی واحد مانند Entity، جایگزینمی کنیم و در نهایت با استفاده از الگویتیم های یادگیری ماشین متون طبقه بندی میشوند. نتایج آزمایشات در این تحقیق، نشان داد که روش دستهبندی k نزدیکترین همسایه به همراه روشهای استخراج ویژگی های لغوی N-gram و کاهش بعد توسط الگوریتم انتخاب ویژگی افزایش اطلاعات، به همراه روش نمونه گیری مجدد از داده ها و حذف موجودیت های موردمقایسه از بردار ویژگی، بیشترین بهبود در کارایی دسته بندها را ایجاد می کند (F\_Measure حدود 81 درصد). بررسی هایانجام شده از مقایسه ی روش پیشنهادی با دسته بندهای پایه در این زمینه، نشان از بهبود 16 درصدی شاخص F\_Measure و موثر بودن مدل پیشنهادی دارد.

## کلمات کلیدی:

نظر کاوی،جملات مقایسه ای،متد افزایش اطلاعات، انتخاب ویژگی،نمونه برداری،طبقه بندی نامتعادل

## لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/783901>



