

عنوان مقاله:

ارائه رویکرد تنسور سه بعدی برای طبقه بندی و تشخیص اخبار جعلی: مطالعه موردی اخبار فارسی در حوزه کرونا ویروس

محل انتشار:

فصلنامه علوم و فنون مدیریت اطلاعات، دوره 7، شماره 4 (سال: 1400)

تعداد صفحات اصل مقاله: 30

نویسندگان:

وحید متقی - دانشجوی دکتری، گروه مدیریت فناوری اطلاعات، واحد قشم، دانشگاه آزاد اسلامی، قشم، ایران.

مهدی اسماعیلی - استادیار، گروه علوم کامپیوتر، واحد کاشان، دانشگاه آزاد اسلامی، کاشان، ایران

قاسمعلی بازاری - استادیار، گروه مدیریت، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران، ایران.

محمدعلی افشارکاظمی - دانشیار، گروه مدیریت، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران، ایران.

خلاصه مقاله:

هدف: هدف پژوهش حاضر اختصاص یکی از کلاس های جعل و واقعی به متن های آزاد می باشد. شبکه های عصبی کانولوشنی به عنوان یکی از مهم ترین مدل های یادگیری عمیق، دقت بالایی را بر روی این مسائل بدست آورده است. در این تحقیق آنالیز متن در سطح جمله و بهبود عملکرد شبکه عصبی کانولوشنی جهت تشخیص اخبار جعلی مورد توجه بوده است. در این شبکه ها کلمات به صورت کیسه ای از کلمات به مدل داده می شوند که هر کلمه با توجه به فضای برداری به ماتریس های دو بعدی تبدیل می شود. یکی از محدودیت های شبکه های کانولوشن این است که در سطح کلمه کار کرده و نمی تواند رابطه و فاصله بین جملات را در نظر بگیرد و آنالیز در سطح جمله مشکل اساسی در این تحقیق می باشد. در این پژوهش یک مدل پایه ای مبتنی بر شبکه های کانولوشنی پیشنهاد شده که در آن اسناد به صورت تنسورهای سه بعدی به شبکه داده می شوند تا بتواند مشکل مذکور را مرتفع نماید. در نظر گرفتن تنسورهای سه بعدی امکان یادگیری موقعیت کلمات در جمله را برای مدل فراهم می آورد و به نتایج دقیق تری در تشخیص اخبار جعل دست می یابد. روش شناسی: پژوهش حاضر مطالعه ای کاربردی بوده که در آن حدود ۴۲۰۰۰ اخبار فارسی از شهرهای مختلف ایران از توییتتر جمع آوری شده و با عمل پیش پردازش، داده های اضافی و غیر مفید حذف و پس از برچسب زدن متون پاک سازی شده، متن اخبار جهت رویکرد پیشنهادی با استفاده از نرم افزار پایتون پردازش شده اند. یافته ها: برخی از الگوریتم های یادگیری ماشین دارای قدرت بیشتری در مسائل طبقه بندی بودند، ولی با تغییراتی که در ساختار الگوریتم شبکه کانولوشن صورت گرفت، نتایج بهتری نسبت به الگوریتم های یادگیری ماشین و سایر الگوریتم های مشابه حاصل شد. نتیجه گیری: در نظر گرفتن تنسورهای سه بعدی امکان یادگیری موقعیت کلمات در جمله را برای مدل فراهم می آورد و این مدل پیشنهادی در مقایسه با رویکردهای پیشنهادی در ادبیات، دقت قابل توجهی را بدست آورده است. مدل پیشنهادی بدون اضافه کردن سربار اضافی از لحاظ تعداد ویژگی ها و عمق شبکه، با تغییر در ورودی توانسته است به نتایج بهتر و قابل قبول از سایر رویکردهای موجود در ادبیات دست یافته و به دقت و صحت بیش از ۹۴ درصد دست یابد.

کلمات کلیدی:

پردازش زبان طبیعی، طبقه بندی متن، شبکه های عصبی کانولوشنی، تنسور سه بعدی، اخبار جعلی، اخبار فارسی، کرونا ویروس

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1429436>



