

عنوان مقاله:

بومی سازی یک الگوریتم ساده سازی متن برای زبان فارسی مبتنی بر یادگیری عمیق و بدون ناظر

محل انتشار:

نخستین همایش ملی هوش مصنوعی و علوم اسلامی (سال: 1399)

تعداد صفحات اصل مقاله: 1

نویسندگان:

ملیکا حامیان - عضو هیئت علمی دانشگاه پیام نور، کارشناسی ارشد مهندسی کامپیوتر گرایش هوش مصنوعی

بهروز جان فدا - دانشجو، کارشناسی ارشد مهندسی کامپیوتر گرایش نرم افزار

سیدعلی حسینی - دکتر؛ پژوهشگر پسادکتری در رشته هوش مصنوعی، گرایش پردازش دانش

بهروز مینایی بیدگلی - دانشیار، (دانشگاه علم و صنعت ایران) دکتری هوش مصنوعی

خلاصه مقاله:

در پردازش زبان طبیعی، ساده سازی متن عملیاتی است که با انجام تغییراتی در کلمات و ساختار زبانی، متن ورودی را از لحاظ خوانایی و فهم پذیری دشوار است، به متنی ساده با خوانایی بالا تبدیل می شود. متون فقهی فارسی و عربی عمدتاً پیچیده هستند. لذا ابزارهای پردازش زبان روی این نوع از متون کارایی چندانی ندارند. برای رفع این مشکل نیاز به الگوریتم های ساده سازی متن داریم. اکثر الگوریتم های موجود در مرزهای دانش در زبان انگلیسی به صورت ناظر آموزش می بینند. یکی از مشکلات عمده در آنها فراهم کردن مجموعه دادگان طلایی است که تهیه آن زمان گزیننده بر است. در حال حاضر چنین دادگانی برای زبان فارسی و عربی موجود نیست، لذا بومی سازی روش های بدون ناظر برای زبان فارسی و عربی منطقی تر به نظر می رسد. در این مقاله ما با بومی سازی یک روش ساده سازی متن بدون ناظر روزآمد برای زبان فارسی، برای اولین بار عملیات ساده سازی متن در زبان فارسی را اجرا می کنیم. روش استفاده شده، برای آموزش نیاز به یک پیکره بزرگ دارد که در آن بدون دخالت انسانی و با کمک یک معیار ریاضی ساده، جملات سخت و ساده برچسب گذاری شده باشند. به این منظور، معیار مورد نظر را برای زبان فارسی تعریف می کنیم و پیکره مورد نظر را با استفاده از پیکره دادگان فارسی می سازیم. روش استفاده شده، یک روش شبکه عصبی عمیق بدون ناظر است و از یک رمزگذار اشتراکی و یکجفت رمزگشا ساخته شده که دانش ساده سازی را از طریق اتلاف های مبتنی بر طبقه بند و متخاصم بدست می آورند. این روش را با استفاده از پیکره ساخته شده آموزش می دهیم و به ساده سازی متون پرداخته و نتایج را گزارش می کنیم. برای ارزیابی عملکرد الگوریتم با معیارها متداول سنجش الگوریتم های ساده سازی متن، وجود یک پیکره طلایی کوچک لازم است که ما برای اولین بار این پیکره آزمایشی کوچک را برای این کار ساخته ایم.

کلمات کلیدی:

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1347048>

