



دانشگاه کاشان
University of Kashan

مجله محاسبات نرم

SOFT COMPUTING JOURNAL

تارنمای مجله: scj.kashanu.ac.ir



طبقه‌بندی متون فارسی مبتنی بر شبکه‌های عصبی عمیق[✦]

محمد رضا فیضی درخشی^۱، استاد، زینب متقی‌نیا^{۲*}، دانشجوی دکتری، میثم عسگری چناقلو^۱، دکتری
^۱ دانشکده مهندسی برق و کامپیوتر، دانشگاه تبریز، تبریز، ایران.

چکیده

امروزه با توجه به رشد روزافزون حجم اسناد الکترونیکی، طبقه‌بندی متون بر اساس روش‌های مختلفی در میان محققین بازایی اطلاعات و متون عمومیت یافته است. با توجه به اهمیت موضوع و کارهایی که در این زمینه در زبان‌های مختلف دنیا انجام گرفته است، نیاز به طبقه‌بندی متون فارسی به خوبی احساس می‌شود. به طور کلی روش‌های طبقه‌بندی متون را می‌توان به روش‌های سنتی (مبتنی بر انتخاب ویژگی و یادگیری ماشین) و روش‌های مبتنی بر یادگیری عمیق تقسیم‌بندی کرد. روش‌های مبتنی بر یادگیری عمیق به دلیل توانایی اشتراک وزن به طور قابل توجهی سبب کاهش تعداد متغیرهای آزاد آموزش‌پذیر شبکه و در نتیجه افزایش تعمیم‌پذیری شده است و نتایج بهتری به نسبت سایر روش‌ها می‌دهد. در زبان فارسی روش‌های مبتنی بر یادگیری عمیق بسیار اندکی برای طبقه‌بندی متون ارائه شده است. در این مقاله دو مدل شبکه عصبی عمیق شامل شبکه عصبی پیچشی ParsCNN و شبکه عصبی با حافظه بلند کوتاه-مدت دوسویه سلسه‌مراتبی با لایه توجه ParsBiLSTM برای طبقه‌بندی متون فارسی تشریح شده است. کارایی سیستم‌های مبتنی بر شبکه عصبی عمیق بر روی مجموعه داده همشهری بررسی شده و از نظر سه معیار ارزیابی دقت، فراخوانی و مقیاس-F مورد مطالعه قرار گرفته است. نتایج آزمایش‌ها نشان می‌دهد که روش ParsCNN میزان دقت ۰/۶۹، فراخوانی ۰/۷ و مقیاس-F ۰/۶۹؛ همچنین روش ParsBiLSTM میزان دقت ۰/۷۲، فراخوانی ۰/۷۳ و مقیاس-F ۰/۷۲ دارند که نشان‌دهنده کارایی بالاتر این روش‌ها نسبت به روش‌های طبقه‌بندی متون فارسی مورد مطالعه است.

© ۱۴۰۱ - مجله محاسبات نرم، کلیه حقوق محفوظ است.

اطلاعات مقاله

تاریخچه مقاله:

دریافت ۲۵ مهر ماه ۱۴۰۰

پذیرش ۲۲ آذر ماه ۱۴۰۱

کلمات کلیدی:

طبقه‌بندی متون

شبکه‌های عصبی عمیق

بردار نمایش کلمات

تعبیه کلمات

یادگیری ماشین

۱. مقدمه

در سال‌های اخیر، تولید حجم عظیمی از داده‌ها سبب شده است تا روش‌های هوشمند استخراج دانش از این مجموعه داده‌ها،

همچون طبقه‌بندی^۱، خوشه‌بندی^۲ [۱]، پیش‌بینی [۲]، ردیابی الگو و ... مورد توجه پژوهشگران قرار بگیرد. طبقه‌بندی متون یکی از موضوعات مطرح در حوزه پردازش زبان طبیعی^۳ [۳، ۴] می‌باشد که وظیفه آن نسبت دادن یک متن به یک یا چند کلاس از پیش تعریف شده است. با توجه به افزایش چشمگیر حجم مطالب متنی در سال‌های اخیر و با توجه به کاربردهای مهم

✦ نوع مقاله: پژوهشی

* نویسنده مسئول

پست(های) الکترونیک: mfeizi@tabrizu.ac.ir (فیضی درخشی)

z.mottagy@gmail.com (متقی‌نیا)

m.asgari.c@gmail.com (عسگری چناقلو)

¹ Classification

² Clustering

³ Natural Language Processing (NLP)