

چکیده:

زبان عربی یک زبان جذاب و محبوب در سطح جهان است. با این وجود، پردازش زبان طبیعی، برای عربی هنوز مورد توجه کافی محققان قرار نگرفته است. یکی از جنبه های مهم این پردازش، تشخیص diacritic کلمات و حروف در زبان عربی است. در زبان عربی، تغییر در diacritic های یک حرف میتواند معنی یک کلمه را به طور کلی عوض کند. به خصوص این یک مشکل بزرگ برای غیر عرب زبانان است. در این تحقیق، یک شبکه مبتنی بر یادگیری عمیق Bi-LSTM ، و CRF با ورودی های دو سطحی ارائه شده است. نتایج تشخیص این الگوریتم سپس با ایجاد برخی پس پردازش ها و اصلاحات، بهتر می شود. نتایج شبیه سازی نشان میدهد الگوریتم پیشنهاد شده در این تحقیق میتواند تا حدود دو برابر از بهترین روش های موجود در این زمینه، بهتر عمل کرده و diacritic های موجود در عبارت عربی را تشخیص دهد.

کلیدواژه ها: پردازش زبان طبیعی ، یادگیری عمیق ، زبان عربی ، diacritic ، Bi-LSTM ، CRF