

چکیده:

با گسترش روز افزون حجم داده‌ها و اطلاعات و همچنین گسترش تعاملات بین‌المللی برقراری ارتباط یکی از مهمترین جنبه‌های زندگی امروز است. مشکل عمده‌ای که در این زمینه وجود دارد عدم امکان برقراری ارتباط با استفاده از داده‌های به زبان دیگر است. از اینرو یکی از مهمترین مسائل زندگی امروز ارائه راه‌حلی خودکار جهت ترجمه از یک زبان به زبان دیگر است. ترجمه ماشینی یکی از راه‌هایی است که برای حل این مشکل ارائه شده است و به خاطر اهمیت آن، در سالیان اخیر توجه بسیار زیادی به آن شده است.

ترجمه ماشینی آماری به عنوان یکی از بهترین روش‌ها برای ترجمه از یک زبان به زبان دیگر شناخته می‌شود. برای زبان‌هایی که از لحاظ ساختار دارای شباهت زیادی به یکدیگر هستند خروجی این مترجم بسیار مناسب است. اما برای برخی جفت زبان‌ها مانند زبان‌های انگلیسی و فارسی تفاوت‌های ساختاری میان دو زبان و همچنین عدم وجود پیکره دوزبانه بزرگ باعث شده است که این روش برای ترجمه انگلیسی به فارسی ترجمه‌های مطلوبی را تولید نکند.

در این پایان‌نامه سعی شده است با کمک گرفتن از اطلاعات زبان‌شناسی، تا حد ممکن بر مشکلات این روش برای ترجمه انگلیسی به فارسی فائق آید. جهت انجام این کار ابتدا سعی در کاهش تفاوت ساختاری میان جملات انگلیسی و فارسی شده است. این عمل می‌تواند منجر به ایجاد مدل ترجمه بهتر شود. برای این منظور یکسری قوانین استخراج و بر روی جملات انگلیسی اعمال گردید. این تغییرات منجر به بهبود حدود 17 درصدی در معیار BLEU و حدود 21 درصدی در معیار NIST گردیده است. در ادامه نسبت به غنی‌سازی عبارات داخل پیکره با استفاده از برخی اطلاعات زبان‌شناسی از جمله برچسب‌های بخش‌های سخن¹ و ریشه کلمات اقدام شد. با این اطلاعات یک سیستم مترجم مبتنی بر فاکتور ایجاد گردید. خروجی این سیستم بهبود حدود 17 درصدی در معیار

¹ Part Of Speech

BLEU و حدود 25 درصدی در معیار NIST را نشان می‌دهد.

کلمات کلیدی:

ترجمه ماشینی، ترجمه ماشینی آماری، مدل زبانی، مدل ترجمه، بازآرایی نحوی، مدلسازی بر مبنای فاکتور