

## عنوان مقاله:

بهبود ترجمه ماشینی آماری با استفاده از دانش زبانی کم عمق

## محل انتشار:

اولین کنفرانس بین المللی روانشناسی، علوم تربیتی و علوم اجتماعی در علوم انسانی (سال: 1403)

تعداد صفحات اصل مقاله: 5

## نویسنده:

ویدا صالحی - دانشجوی زبان انگلیسی مقطع کارشناسی

## خلاصه مقاله:

ما روش هایی را برای بهبود عملکرد ترجمه ماشینی آماری (SMT) بین چهار زبان مختلف از نظر زبانی، یعنی چینی، انگلیسی، ژاپنی و کره ای با استفاده از دانش مورفوسنتکسی توصیف می کنیم. به منظور کاهش ابهامات ترجمه و تولید خروجی ترجمه صحیح و روان از نظر دستوری، از دانش زبانی کم عمق استفاده می کنیم، یعنی: (۱) غنی سازی یک کلمه با ویژگی های صرفی آن، (۲) به دست آوردن جفت عبارات پراکنده. عمق با انگیزه زبانی، (۳) تراز تکراری کلمات با استفاده از جفت عبارات فیلتر شده و (۴) ساختن یک مدل زبانی از کلمات غنی شده از لحاظ مورفوسنتکسی. این مقاله یک مدل سلسله مراتبی جدید برای ترجمه ماشینی آماری پیشنهاد می کند که غیر پایانه ها را با تطبیق مرزهای عبارت هدف با برچسب های کم عمق نحوی در سمت هدف مجموعه آموزشی نام گذاری می کند. در جایی که هیچ برچسبی برای کل عبارت وجود ندارد، نام غیر پایانی با اضافه کردن برچسب های مرزی تعریف می شود. برچسب گذاری کلاس کلمه مرزهای عبارت قبلا معرفی شده است که می تواند مبنای مدل پیشنهادی باشد. ما این فرم اصلی را در مقاله حاضر با استفاده از برچسب های قطعه توسعه می دهیم. در این توسعه، اگر تگ قطعه در عبارت مرزی وجود نداشته باشد، از پیشوند POS کلمه مرزی استفاده می شود. با استفاده از برچسب های عبارت به جای کلاس های کلمه، قوانین مدل پیشنهادی تعمیم می یابد. تعدادی آزمایش به صورت ترجمه فارسی به انگلیسی انجام شد. با استفاده از معیار BLEU، در مقایسه با مدل SAMT که از درخت تجزیه برای برچسب گذاری استفاده می کند، مدل پیشنهادی به بهبود قابل توجهی دست می یابد.

## کلمات کلیدی:

ترجمه ماشینی آماری - مدل زنجیره ای - خلاصه کلمه

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/2025898>

