

5 - نتیجه گیری و کارهای آتی

5-1 - نتیجه گیری

در این پایان نامه ابتدا یک سیستم مترجم آماری انگلیسی به فارسی مبتنی بر عبارت ارائه شده است. از آنجایی که یکی از معضلات موجود در ساخت چنین سیستمی نبود پیکره بزرگ و مناسب است، دو روش که در آن ها استفاده صریح از اطلاعات زبان شناسی مطرح شده است، جهت بهبود خروجی مترجم ماشینی آماری انگلیسی به فارسی پیشنهاد شده است. در روش اول به کاهش فاصله ساختاری بین زبان انگلیسی و فارسی پرداخته شده است. این عمل باعث ترازبندی بهتر عبارات معادل در دو زبان و در نتیجه ایجاد مدل ترجمه بهتر می گردد. خروجی سیستم بر روی پیکره ای شامل حدود یازده هزار جمله تراز شده تحت دو معیار BLEU و NIST و برای دو مجموعه تست متفاوت استخراج گردیده است. مقایسه نتایج نشان دهنده بهبود روش پیشنهادی در نتایج نهایی ترجمه می باشد.

در ادامه به بررسی تاثیر فاکتورهای زبانی در مراحل ترجمه پرداخته شده است. برای این منظور اثر اضافه نمودن فاکتورهای ریشه کلمات و برچسب های بخش های سخن درون پیکره موازی بررسی شده است. در این مدل ترجمه مبتنی بر فاکتور پیاده سازی گردیده است. در این روش ریشه کلمات و برچسب های بخش های سخن به کمک ترازبندی آمده و منجر به بهبود مدل ترجمه می گردند. همچنین با استفاده از ریشه کلمات موجود در پیکره تا حدی مشکل ترجمه برخی کلمات خارج از پیکره نیز مرتفع می گردد. خروجی حاصل از اعمال این روش باز هم منجر به بهبود معیارهای BLEU و NIST نسبت به سیستم پایه گردیده است.

5-2 - کارهای آتی

یکی از مسائل مهم در زمینه روش های بهبود سیستم های مترجم ماشینی آماری، بررسی میزان بهبود

سیستم به ازای اندازه‌های مختلف پیکره است. از آنجایی که حجم پیکره در این پایان‌نامه به قدری نبود که بتوانیم این آزمایشات را انجام دهیم می‌توان به عنوان کارهای آتی به ایجاد یک پیکره نسبتاً بزرگ پرداخت و اثر روش‌های پیشنهاد شده را بر روی اندازه‌های متفاوت آزمایش نمود. البته بدیهی است که اگر اندازه پیکره خیلی بزرگ باشد میزان اثر این روش‌ها کاهش خواهد یافت.

یکی دیگر از نکاتی که مهم به نظر می‌رسد بررسی سایر فاکتورهای زبانی از جمله جنسیت¹، تعداد²، نقش‌های معنایی کلمات و سایر تحلیل‌های صرفی موجود است. همچنین با بالا بردن دقت ابزار تولید شده در این پایان‌نامه می‌توان در جهت بهبود نتایج نهایی در ترجمه گام برداشت.

نکته دیگری که می‌توان روی آن تمرکز نمود ارائه ابزار تجزیه‌گر برای زبان فارسی است. با انجام این کار می‌توان عمل بازآرایی را برای زبان فارسی انجام داد و سیستم مترجم آماری فارسی به انگلیسی را بهبود بخشید. از آنجایی که زبان انگلیسی از لحاظ ترتیب اجزاء قانونمندتر از زبان فارسی است این عمل می‌تواند بهبود مناسبی در نتایج نهایی ترجمه ایجاد نماید.

¹ gender
² count