

## ■ دفاع از رساله دکتری

□ سمینار عمومی (Colloquium)

□ دفاع از پایان نامه کارشناسی ارشد

□ سمینار تخصصی (Seminar)

□ سمینار تخصصی و مشورتی (Informal Seminar)

## عنوان: ارائه یک مدل ترجمه ماشینی آماری مبتنی بر الگوی ترازبندی سلسله مراتبی

سخنران: شهرام سلامی

## چکیده:

ترجمه ماشینی یکی از شاخه‌های پرکاربرد پردازش زبان طبیعی است که با مطرح شدن روش‌های ترجمه آماری در دهه اخیر استفاده از آن فراگیر شده است. مدل‌های ترجمه آماری در دو رده مبتنی بر عبارت و سلسله مراتبی دسته بندی می‌شوند. مدل‌های سلسله مراتبی برای انجام بهتر بازترتیب کلمات در جفت زبان‌هایی با تفاوت زیاد در ترتیب کلمات مطرح شده‌اند. از سوی دیگر، تفاوت در ترتیب کلمات زبان‌ها موجب کاهش تعداد عبارات تراز شده و افزایش تنگی مدل ترجمه می‌گردد.

در این رساله مدل جدیدی با نام مدل عبارت-مرزی در رده مدل‌های سلسله مراتبی مطرح می‌گردد. برای کاهش تنگی در این مدل، تنها کلمات مرزی عبارات در برجسب گذاری قواعد شرکت می‌کنند. مدل عبارت-مرزی بر این فرض استوار است که کلمات مرزی عبارات، نقش مهمی در امکان هم‌جواری آنها در جمله دارند. برجسب قواعد در مدل پیشنهادی با برجسب‌های مرزی سمت مقصد عبارات تراز شده تعریف می‌شود. میزان اطلاعات زبانی مورد استفاده به نوع برجسب‌های پیکره مقصد بستگی دارد. برجسب‌ها در شکل مبنای مدل با کلاس کلمات تعریف می‌شود که با خوشه بندی کلمات پیکره مقصد یا برجسب POS آنها قابل تعیین است. در توسعه مدل عبارت-مرزی از اطلاعات زبانی بیشتری استفاده می‌شود و برجسب قواعد با استفاده از برجسب‌های کم عمق زبانی در مرز عبارات شامل برجسب POS، برجسب قطعه و یا نقش معنایی تعریف می‌شود. لازم به ذکر است که علاوه بر امکان خوشه بندی کلمات، امکان تعریف برجسب POS و قطعات برای بیشتر زبان‌ها از جمله زبان‌هایی با کمبود منابع وجود دارد.

اندازه مدل و زمان ترجمه چالش‌های مهمی در ترجمه ماشینی آماری و به خصوص در مدل‌های سلسله مراتبی است. این رساله برای حفظ کارایی مدل عبارت-مرزی، روش جدیدی را برای فیلتر قواعد بر مبنای الگوی تراز عبارات مطرح می‌کند. در این روش با توجه به نوع تجزیه پذیری عبارات تراز شده، دو فیلتر غیروازگانی و یکنوا پیشنهاد می‌گردد. فیلتر غیروازگانی به قواعد انتزاعی با غیرپایانه‌های ریز دانه نیاز دارد که برای شکل مبنای مدل عبارت-مرزی با تعداد زیاد کلاس کلمات قابل استفاده است. فیلتر یکنوا قواعد استخراج شده از عبارات با تراز یکنوا را کاهش می‌دهد و برای همه مدل‌های سلسله مراتبی کاربرد دارد. علاوه بر کاهش اندازه مدل و زمان ترجمه، این فیلترها منابع لازم برای تولید گرامر را نیز کاهش می‌دهند.

مجموعه‌ای از آزمایش‌ها در ترجمه از فارسی، آلمانی و فرانسه به انگلیسی به عنوان جفت زبان‌هایی با تفاوت زیاد، متوسط و کم در ترتیب کلمات انجام شد. با استفاده از معیار BLEU در این آزمایش‌ها، مدل عبارت-مرزی با مدل مبتنی بر عبارت سلسله مراتبی و مدل SAMT که از گره‌های درخت تجزیه نحوی برای برجسب گذاری استفاده می‌کند، مقایسه گردید. کیفیت ترجمه شکل مبنای مدل عبارت-مرزی با SAMT مشابه و از مدل مبتنی بر عبارت سلسله مراتبی بالاتر است. زمان ترجمه این مدل‌ها نیز نسبت به مدل مبتنی بر عبارت سلسله مراتبی بسیار بیشتر است که با فیلتر قواعد تا حد زیادی کاهش می‌یابد. در ترجمه از فارسی و آلمانی به انگلیسی به عنوان جفت زبان‌هایی با ترتیب غیر یکنوا کلمات، توسعه مدل عبارت-مرزی، کیفیت ترجمه را به میزان قابل توجهی افزایش می‌دهد.

کلمات کلیدی:

زمان برگزاری: ۱۳۹۶/۶/۶

مکان برگزاری: دانشکده علوم و مهندسی کامپیوتر