تخمین اطمینان خروجی ترجمهٔ ماشینی با استفاده از ویژگی‌های جدید ساختاری و محتوایی

مرضیه صالحی[1]، شهرام خدیوی[2] و نوشین ریاحی[3]

[1و3] دانشکدهٔ مهندسی، دانشگاه الزهرا (س)، تهران، ایران

[2] آزمایشگاه تکنولوژی زبان انسان، دانشکدهٔ مهندسی کامپیوتر، دانشگاه امیرکبیر، تهران، ایران

چکیده

با وجود پیشرفت‌های اخیر در حوزهٔ ترجمهٔ ماشینی، این فناوری قادر به ترجمهٔ دقیق متون نیست و گاهی ممکن است ویرایش خروجی آن زمان بیشتری نسبت به ترجمهٔ دستی بگیرد. با این حال با داشتن تخمینی از کیفیت خروجی، کاربران می‌توانند به‌طور مناسبی با ناکامل‌بودن این فناوری برخورد کنند. برای کاربردهایی که هدف آنها بالابردن کیفیت ترجمهٔ ماشینی است، نظیر ترکیب خروجی سامانه‌های ترجمه‌گر مختلف، بازترتیب فهرست چند ترجمهٔ بهتر و بازتولید خروجی، لازم است که بدون داشتن ترجمهٔ مرجع تخمینی از درستی خروجی داشته باشیم. هنوز روش کارامدی برای تخمین درستی کلمات خروجی ترجمه ماشینی وجود ندارد. در این مقاله پنج گروه ویژگی جدید در قالب ویژگی‌های مبتنی بر محتوا و مبتنی بر ساختار ارائه شده است. نتایج نشان می‌دهد که ویژگی مبتنی بر محتوا نسبت به بهترین سامانهٔ پایه (۲) برتری ۹/۶۳ درصدی در CER، ۸/۵ درصدی در F-measure و ۵/۱ درصدی در F-measure طبقهٔ منفی داشته است. همچنین ترکیب ویژگی‌های مبتنی بر ساختار ارائه‌شده، در مقایسه با بهترین سامانهٔ پایه به‌ترتیب بهبود ۴/۵۹، ۴/۱ و ۲ درصدی در معیارهای CER، F-measure و F-measure طبقه منفی ایجاد کرده است.

واژگان کلیدی: تخمین اطمینان، ترجمه ماشینی، اطلاعات متقابل، ساختار، محتوا.

۱- مقدّمه

با وجود اینکه از زمان ظهور ترجمهٔ ماشینی یعنی درحدود شصت سال پیش، بهبودهای وسیعی در این حوزه به وجود آمده است؛ کارایی سامانه‌های ترجمهٔ ماشینی امروزی هنوز فاصلهٔ زیادی با کامل‌شدن دارد (ریبود و همکاران، ۲۰۱۱). یکی از دلایل این مسئله ابهام ذاتی زبان طبیعی است؛ به‌طوری‌که همهٔ فناوری‌های جاری مربوط به زبان طبیعی خطا دارند (بلاتز و همکاران، ۲۰۰۴). روش‌های آماری در دو دههٔ اخیر کاربرد گسترده‌ای در تشخیص گفتار، پردازش زبان طبیعی و ترجمهٔ ماشینی داشته و به موفقیت‌های قابل ملاحظه‌ای رسیده‌اند. این موفقیت‌ها به‌دلیل رویکرد مستقل از زبان و عدم نیاز به دانش اولیه است. در این روش‌ها تنها به پیکره‌های به اندازهٔ کافی بزرگ برای تخمین چگالی‌های احتمال نیاز است؛ اما روش‌های آماری همیشه از منظر انسان بهینه نیستند؛ (ریبود و همکاران، ۲۰۰۹ب) بلکه خود منشأ ایجاد خطا هستند. با این حال با یک تخمین از کیفیت خروجی فناوری‌های ناکامل، کاربران می‌توانند با ناکامل‌بودن این فناوری‌ها برخورد مناسبی کنند. برای مثال گاهی ممکن است ویرایش ترجمهٔ تولیدشده توسط ماشین زمان بیشتری نسبت به ترجمهٔ دستی بگیرد. در چنین شرایطی مطلوب است ویرایش‌گران تخمینی از درستی بخش‌های مختلف ترجمه داشته باشند تا تلاش خود را معطوف به بخش‌هایی کنند که از یک سو به‌طور کامل درست نیستند و از سوی دیگر نیاز به تغییرات پرهزینه ندارند (بلاتز و همکاران، ۲۰۰۴).

علاوه‌بر پس‌ویرایش، تخمین اطمینان برای کاربردهایی که هدف آنها بالابردن کیفیت ترجمهٔ ماشینی است نیز مفید می‌باشد. از جملهٔ این کاربردها می‌توان به ترکیب خروجی چند سامانهٔ ترجمه‌گر هیلدربرند و ووگل، ۲۰۰۸)، مرتب‌سازی دوباره فهرست چند ترجمهٔ بهتر (بچ و همکاران، ۲۰۱۱) و بازتولید فرضیات جدید، با ترکیب دوباره

بخش‌هایی از ترجمه‌های نامزد که درجهٔ اطمینان بالاتری دارند، اشاره کرد. همچنین تخمین اطمینان خروجی ترجمه ماشینی در ترجمهٔ ماشینی آماری تعاملی کاربرد دارد (اوفینگ و نای، ۲۰۰۵).

هرچند پیشینهٔ کاربرد تخمین اطمینان به شناسایی گفتار می‌رسد (اوفینگ و نای، ۲۰۰۵)،، اما تاکنون تلاش‌های متعددی در زمینهٔ بهبود تخمین اطمینان ترجمهٔ ماشینی صورت گرفته است. با وجود این، هنوز دقت آن به حد مناسبی نرسیده است؛ به‌طوری‌که برای مثال جواب‌گوی کاربرد پس‌ویرایش نیست (ریبود و همکاران، ۲۰۱۱).

تخمین اطمینان، رویکرد یادگیری ماشینی دارد و از سه مرحلهٔ اصلی تشکیل می‌شود: ۱. محاسبهٔ ویژگی‌ها، ۲. ترکیب ویژگی‌ها با استفاده از یک مدل طبقه‌بندی مناسب و ۳. طبقه‌بندی کلمات با استفاده از حد آستانه بهینه‌شده برای طبقه‌بندی. گاهی مراحل دوم و سوم با هم ادغام شده و در یک مرحله انجام می‌شوند (زیونگ و همکاران، ۲۰۱۰). در برخی از موارد فقط یک ویژگی استخراج و طبقه‌بندی با انتخاب حد آستانه انجام می‌شود که در این صورت به آن ویژگی، درجهٔ اطمینان می‌گویند.

تخمین اطمینان را می‌توان از جهت نوع خروجی به دو دسته تقسیم کرد: قوی و ضعیف (بلاتز و همکاران، ۲۰۰۴). تخمین اطمینان قوی احتمال درستی یا امتیاز درستی خروجی ترجمهٔ ماشینی را برمی‌گرداند و برای انجام آن از روش‌های رگرسیون، از جمله رگرسیون کمینهٔ مربعات جزیی (اسپشیا و همکاران، ۲۰۰۹)، رگرسیون منطقی (ریبود و همکاران، ۲۰۰۹ب) و ماشین بردار پشتیبان (ریبود و اسمایلی، ۲۰۰۹) استفاده می‌شود. در رویکرد ضعیف به خروجی ترجمهٔ ماشینی برچسب زده می‌شود. این برچسب‌ها می‌توانند به‌صورت {"خوب"، "بد"} (اوفینگ و نای، ۲۰۰۳؛ بلاتز و همکارن، ۲۰۰۴؛ ریبود و همکاران، ۲۰۰۹الف؛ ریبود و همکاران، ۲۰۰۹ب، زیونگ و همکاران، ۲۰۱۰) و یا {"افزوده شده"، "جابه‌جا شده"، "جانشین شده"، "خوب"} (بچ و همکاران، ۲۰۱۱) باشند. در تخمین اطمینان ضعیف به‌طور معمول از روش‌های طبقه‌بندی نظیر نایوبیز (بلاتز و همکاران، ۲۰۰۴)، پرسپترون چندلایه با یک لایهٔ پنهان (بلاتز و همکاران، ۲۰۰۴) با تعداد متفاوت لایهٔ پنهان (کویرک، ۲۰۰۴)، ماشین بردار پشتیبان (کویرک، ۲۰۰۴)، بیشینهٔ آنتروپی (زیونگ و همکارن، ۲۰۱۰) و درخت تصمیم (کویرک، ۲۰۰۴) استفاده می‌شود.

از نگاه واحد مورد آزمون، می‌توان تخمین اطمینان را به سه سطح تقسیم کرد: سطح کلمه، سطح زیرجمله و سطح جمله. آنچه در این مقاله به آن پرداخته شده، تخمین اطمینان سطح کلمه است. تاکنون برای تخمین اطمینان سطح کلمه ویژگی‌های متعددی پیشنهاد شده است. این ویژگی‌ها را می‌توان با نگاه تحلیلی به‌صورت زیر دسته‌بندی کرد:

۱-۱- ویژگی‌های پیش‌ترجمه

ویژگی‌هایی که از اطلاعات مربوط به کلمهٔ مبدأ، زمینهٔ[1] کلمهٔ مبدأ و سامانهٔ ترجمه‌گر استفاده می‌کنند. از جمله ویژگی‌های ارائه‌شده که در این دسته قرار می‌گیرند، می‌توان به ویژگی‌های سمت مبدأ (بچ و همکاران، ۲۰۱۱) و بلندترین طول مدل زبانی مبدأ (لونگ، ۲۰۱۳) اشاره کرد. درواقع ویژگی‌های این دسته، پیش از انجام فرآیند ترجمه با توجه به کلمهٔ مبدأ، زمینهٔ آن و سامانهٔ ترجمه‌گر، تشخیص می‌دهند که کلمهٔ مورد نظر درست یا نادرست ترجمه خواهد شد. سامانه‌های تخمین اطمینانی که از این ویژگی‌ها استفاده می‌کنند، به‌طور معمول وابسته به سامانهٔ ترجمه‌گر هستند. برای مثال به‌سختی می‌توان ویژگی مربوط به سمت مبدأیی ارائه داد و تضمین کرد که همهٔ سامانه‌های ترجمه‌گر برای مقادیر مختلف آن، از نظر صحت ترجمه، یکسان عمل می‌کنند. بنابراین سامانه‌ای که از این ویژگی‌ها استفاده می‌کند، برای سامانه‌های ترجمهٔ مختلف، متفاوت عمل می‌کند.

۱-۲- ویژگی‌های پس‌ترجمه

اطلاعاتی که این ویژگی‌ها از آن استفاده می‌کنند، شامل اطلاعات مربوط به ترجمه‌های تولیدشده است. ویژگی‌هایی که در این گروه قرار می‌گیرند، باید بعد از انجام فرآیند ترجمه و یا درحین آن محاسبه شوند. این گروه خود به دو گروه ویژگی‌های مبتنی بر سامانهٔ ترجمه‌گر و ویژگی‌های مستقل دسته‌بندی می‌شود.

۱-۲-۱- ویژگی‌های مبتنی بر سامانهٔ ترجمه‌گر

این ویژگی‌ها ویژگی‌های جعبهٔ شیشه‌ای[2] (اسپشیا و همکاران، ۲۰۰۹) و ویژگی‌های موتور[3] (باک، ۲۰۱۲) نیز

[1] Context
[2] Glass box features
[3] Engine features

نامیده شده‌اند. ویژگی‌های این دسته به‌طور معمول به کار داخلی سامانه ترجمه‌گر مربوط هستند و از خروجی‌های مربوط به حین ترجمه مانند فهرست چند ترجمهٔ بهتر[1] یا گراف کلمات استفاده می‌کنند. از جمله ویژگی‌های ارائه‌شده در این دسته می‌توان به احتمال پسین، جمع رتبه‌ها و فرکانس مربوط (اوفینگ و نای، ۲۰۰۳) اشاره کرد.

۱-۲-۲- ویژگی‌های مستقل

دستهٔ دیگر ویژگی‌های گروه پس‌ترجمه از اطلاعات مربوط به کلمه فرضیه (مقصد) و میزان سازگای آن با جملات مبدأ و مقصد استفاده می‌کنند. این ویژگی‌ها مستقل از سامانهٔ ترجمه‌گر هستند. این گروه به دو گروه مبتنی بر ساختار و مبتنی بر محتوا دسته‌بندی می‌شود.

ویژگی‌های مبتنی بر ساختار: این ویژگی‌ها سازگاری صرفی یا نحوی کلمهٔ مقصد را با جمله مقصد و یا مبدأ بررسی می‌کنند. از جمله ویژگی‌های ارائه‌شده که در این دسته قرار می‌گیرند ساختار وابستگی مبدأ و مقصد (بچ و همکاران، ۲۰۱۱) الگوی برچسب‌های جزء کلام کلمات مقصد (زیونگ و همکاران، ۲۰۱۰)، وابستگی‌های تهی (زیونگ و همکاران، ۲۰۱۰)، برچسب کلمه و عمق آن در درخت (لونگ و همکاران، ۲۰۱۳)، ویژگی‌های مبتنی بر مدل زبانی برچسب جزء کلام (لونگ و همکاران، ۲۰۱۴) و احتمال چندتایی شخص، زمان و جنسیت کلمات (ریبود و همکاران، ۲۰۰۹الف) می‌باشند.

ویژگی‌های مبتنی بر محتوا: این ویژگی‌ها سازگاری لغوی یا معنایی کلمهٔ مقصد را با جملهٔ مقصد و یا با جملهٔ مبدأ بررسی می‌کنند. از ویژگی‌های این گروه می‌توان ویژگی‌های مبتنی بر اطلاعات متقابل (ریبود و همکاران، ۲۰۰۹الف؛ ریبود و همکاران، ۲۰۰۹ب)، رخداد کلمه در چندین ترجمه (لونگ و همکاران، ۲۰۱۴)، ویژگی‌های مبتنی بر مدل IBM-1 و شباهت معنایی متوسط (بلاتز و همکاران، ۲۰۰۴) را نام برد.

در این مقاله به‌منظور داشتن سامانهٔ تخمین اطمینانی مستقل از سامانهٔ ترجمه‌گر، تمرکز بر روی ویژگی‌های مستقل است. با وجود اینکه تلاش‌های گذشته در زمینهٔ ویژگی‌های مستقل موجب پیشرفت‌هایی شده است، این ویژگی‌ها معایبی نیز دارند. به‌جز ویژگی ساختار وابستگی مبدأ و مقصد (بچ و همکاران، ۲۰۱۱) ویژگی‌های مبتنی بر ساختار بیش‌تر سازگاری کلمهٔ مقصد را با زمینهٔ آن در جملهٔ مقصد بررسی می‌کنند. از طرفی اکثر جملات مقصد دارای خطا هستند. در (بلاتز و همکاران، ۲۰۰۴) سنجش کیفیت خروجی ترجمهٔ ماشینی توسط انسان انجام و گزارش شد که فقط ۳۰٪ از جملات خروجی امتیاز بالا گرفتند. بخشی از این خطاها خطای ساختاری است و به‌ندرت جملات خروجی ترجمه‌گر ماشینی از نظر ساختاری کاملند (زیونگ و همکاران، ۲۰۱۰). در این صورت خطاهای ساختاری موجود در جملهٔ مقصد که توسط سامانهٔ ترجمه‌گر تولید شده است، موجب بروز خطا در تصمیم‌گیری در مورد صحت کلمهٔ مقصد مورد آزمون می‌شود. همچنین برای اکثر زبان‌ها، تجزیه‌گر نحوی‌ای که بتواند جملاتی را که از نظر ساختاری ناقص هستند، با دقت قابل قبولی تجزیه کند، وجود ندارد. برای حل این مشکلات، در این مقاله از جملهٔ مبدأ که ساختار کاملی دارد، برای بررسی صحت کلمهٔ مقصد استفاده شده است. به بیان دیگر در ویژگی‌های ارائه‌شده در این مقاله، سازگاری ساختاری کلمات مقصد با جملات مبدأ مورد بررسی قرار گرفته است. درنتیجه این ویژگی‌ها تنها نیاز به تجزیهٔ نحوی جملات مبدأ دارند. در زمینهٔ ویژگی‌های مبتنی بر محتوا نیز تاکنون پژوهش‌های زیادی صورت نگرفته است. با این حال نتایج پژوهش‌های انجام‌شده نشان می‌دهد در میان ویژگی‌های ارائه‌شده، ویژگی اطلاعات متقابل بین زبانی (ریبود و همکاران، ۲۰۰۹الف) عملکرد خوبی نسبت به اکثر ویژگی‌های دیگر دارد. در این مقاله با ایجاد تغییرات منطقی بر روی این ویژگی، کارایی آن‌را به‌طور قابل توجهی بهبود داده‌ایم. یکی دیگر از چالش‌های پیش رو برای تخمین اطمینان، تعداد نمونه‌های آموزشی است؛ حتی اگر از روش‌های خودکار برای تولید نمونه‌های آموزشی برچسب‌خورده استفاده شود، از آنجا که پیکرهٔ آموزش، پیش از برچسب‌زنی باید برای ترجمه به‌عنوان پیکرهٔ آزمون به سامانه ترجمه‌گر داده شود، از بخش بزرگی از جملات موازی موجود برای یک جفت زبان (جملات استفاده‌شده برای آموزش سامانهٔ ترجمه‌گر) نمی‌توان به‌عنوان پیکرهٔ آموزش تخمین اطمینان استفاده کرد. برای مقابله با این چالش در این مقاله ویژگی‌هایی ارائه‌شده که دو مرحلهٔ آموزش را طی می‌کنند. برای محاسبهٔ این ویژگی‌ها از چگالی‌های احتمال استفاده می‌شود. در مرحلهٔ اول چگالی‌های احتمال با استفاده از یک پیکرهٔ بزرگ که می‌تواند شامل پیکرهٔ آموزش سامانهٔ ترجمه‌گر نیز باشد، استخراج می‌شوند. در مرحلهٔ دوم پس از محاسبهٔ ویژگی‌ها با استفاده از مشخصه‌های آماری حاصل از مرحلهٔ قبل، سامانهٔ تخمین اطمینان در چارچوب

[1] N-best list

یک روش طبقه‌بندی آموزش می‌بینند. درنهایت برای ترکیب ویژگی‌های ارائه‌شده از روش‌های یادگیری تمایزی متفاوت بهره گرفته شده و نتایج حاصل با هم مقایسه شده است. مزیت روش‌های تمایزی نسبت به روش‌های مولد این است که فرض عدم وابستگی ویژگی‌ها به هم را ندارند.

۲- ویژگی‌ها

در این مقاله مجموعه‌ای از ویژگی‌های مستقل ارائه کرده‌ایم. ویژگی مبتنی بر اطلاعات متقابل بهبودیافته به‌عنوان یک ویژگی مؤثر مبتنی بر محتوا و ویژگی‌های احتمال هم‌ترازی نقش دستوری، سازگاری فعل، عبارات دستوری و جایگاه نسبی کلمه در زیرجمله به‌عنوان ویژگی‌های مبتنی بر ساختار ارائه شده‌اند که در ادامه جزئیات آنها را شرح می‌دهیم.

۲-۱- ویژگی مبتنی بر اطلاعات متقابل

در نظریهٔ احتمال، اطلاعات متقابل میزان وابستگی متقابل دو متغیر تصادفی را اندازه می‌گیرد و می‌توان از آن برای مشخص‌کردن جفت‌کلماتی که محتمل است با هم در جملات ظاهر شوند، استفاده کرد. در (گو و همکاران، ۲۰۰۴) یک درجهٔ اطمینان سطح کلمه بر پایهٔ اطلاعات متقابل برای شناسایی گفتار ارائه شده است. در (ریبود و همکاران، ۲۰۰۹الف) برای تخمین اطمینان ترجمهٔ ماشینی، دو درجهٔ اطمینان مبتنی بر اطلاعات متقابل بین زبانی و میان زبانی ارائه شده است.

۲-۱-۱- درجهٔ اطمینان مبتنی بر اطلاعات متقابل بین زبانی

اگر تعداد جملات مبدأ را که شامل کلمه x هستند با $N_S(x)$ ، تعداد جملات هدف شامل y را با $N_T(y)$ و تعداد جفت جملات مبدأ و هدف به‌ترتیب شامل x و y را با $N(x,y)$ نمایش دهیم، و تعداد کل جفت‌جملات مبدأ و هدف را با N نمایش دهیم، روابط (۱) تا (۴) تعریف می‌شوند:

$$p_t(y)=\frac{N_t(y)}{N} \qquad (۱)$$

$$p_s(x)=\frac{N_s(x)}{N} \qquad (۲)$$

$$p(x,y)=\frac{N(x,y)}{N} \qquad (۳)$$

$$MI(x,y)= p(x,y)\log_2\left(\frac{p(x,y)}{p_s(x)p_t(y)}\right) \qquad (۴)$$

و درجه اطمینان کلمه j ام جملهٔ مقصد (t_j) طبق رابطهٔ (۵) با اعمال یک تابع وزن‌دهی برای ترکیب خطی $MI(s_i,t_j)$ ها به‌ازای s_i های متغیر روی کلمات جمله مبدأ به‌دست می‌آید.

$$C(t_j)=\frac{\sum_{i=1}^{I} w(|j-i|)\, MI(s_i,t_j)}{\sum_{i=1}^{I} w(|j-i|)} \qquad (۵)$$

در (ریبود و همکاران، ۲۰۰۹ب) از پنجرهٔ رهاسازی به‌عنوان تابع وزن‌دهی استفاده شده است که مرکز پنجره جایگاه کلمهٔ مقصد در جملهٔ مقصد یعنی j است؛ یعنی به‌عنوان مثال برای پنجره‌ای با طول ۵، w(.) به‌صورت زیر تعریف می‌شود:

$$w(k)=\begin{cases}1 & k<3\\ 0 & k\geq 3\end{cases} \qquad (۶)$$

این نوع وزن‌دهی برای جفت‌زبان‌هایی که در آنها ترتیب کلمات شبیه هم است (مانند فرانسه-انگلیسی که در (ریبود و همکاران، ۲۰۰۹ب) استفاده شده است.) مناسب است؛ اما برای جفت‌زبان‌هایی که ترتیب کلمات در جمله برای آنها متفاوت است، مانند انگلیسی-فارسی، مناسب نیست. برای حل این مشکل تغییراتی در این ویژگی ایجاد شده است که در بخش ۲-۱-۲ بیان می‌شود.

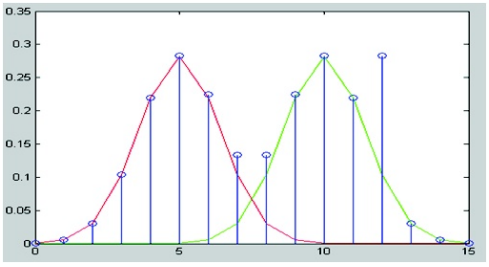
۲-۱-۲- ویژگی مبتنی بر اطلاعات متقابل بین زبانی بهبودیافته

یک کلمه مقصد دلخواه مانند t_j براساس جملهٔ مبدأ تولید می‌شود. اما برخی از کلمات جملهٔ مبدأ نسبت به سایر کلمات مبدأ، ارتباط بیشتری با کلمه t_j دارند. بنابراین انتظار می‌رود این کلمات مبدأ در محاسبهٔ ویژگی اطلاعات متقابل بین زبانی کلمه t_j وزن بیشتری بگیرند. یکی از کلمات مبدأیی که ارتباط زیادی با کلمات مقصد دارد، فعل جملهٔ مبدا است. همچنین انتظار می‌رود کلماتی که سامانهٔ ترجمه‌گر کلمه t_j را ترجمه آنها می‌داند (درصورت مشخص نشدن هم‌ترازی‌ها توسط سامانهٔ ترجمه‌گر، می‌توان از هم‌ترازی IBM-1 استفاده کرد.) وزن بیشتری در محاسبهٔ اطلاعات متقابل بین زبانی بگیرند. در این مقاله به‌منظور عملی‌کردن این انتظارات، برای وزن‌دهی، مدل‌هایی ارائه شده و پارامترهای آنها با استفاده از پیکرهٔ آموزش و توسعه تخمین زده شده است. در ادامه این مدل‌ها توضیح داده می‌شود.

مدل نمایی: اگر کلمات مبدأیی را که کلمه مقصد t_j ترجمه آنها در جمله مورد نظر است، مجموعه $S_a=\{s_i|(i,j)\in A\}$ بنامیم، که A مجموعه هم‌ترازی‌های

موجود بین کلمات مبدأ و مقصد به‌ازای جمله مورد نظر باشد، در مدل وزن‌دهی نمایی، مطابق شکل ۱، حول هرکدام از اعضای S_a به‌عنوان مثال s_i یک تابع وزن‌دهی نمایی در نظر گرفته می‌شود که با دورشدن از s_i وزن کلمات به‌صورت نمایی کاهش می‌یابد. همچنین برای فعل یا افعال جملهٔ مبدأ وزن جداگانه‌ای در نظر گرفته می‌شود. وزن‌دهی نهایی از حاصل جمع وزن‌های حول هر یک از s_i ها به‌دست می‌آید. در شکل ۱ محور افقی شمارهٔ کلمات در جملهٔ مبدأ و محور عمودی مقدار وزن‌ها را نشان می‌دهد. منحنی قرمزرنگ و سبزرنگ هر کدام تابعی نمایی حول دو کلمه هم‌تراز با کلمهٔ مقصد هستند که روی کل جملهٔ مبدأ تعریف می‌شوند. خطوط آبی نشان‌دهندهٔ وزن‌های نهایی اختصاص‌یافته به هر یک از کلمات جملهٔ مبدأ می‌باشند. کلمهٔ فعل (کلمه ۱۲ ام) نیز وزن جداگانه‌ای می‌گیرد. رابطهٔ (۷) نشان‌دهنده تابع وزن‌دهی است.

$$w(k)=\begin{cases}\sum_{i|(i,j)\in A} e^{\frac{(k-i)^2}{\delta^2}} & s_k \text{ is not verb}\\ f(w(k)) & s_k \text{ is verb}\end{cases} \qquad (۷)$$



(شکل ۱): وزن‌دهی در مدل نمایی

f(.) تابعی خطی از مقادیر w(k) است. درواقع چنین فرض شده که در هر جمله سهم مشخصی از کل وزن‌ها به فعل جمله اختصاص می‌یابد. مقدار بهینهٔ پارامتر δ و تابع f(.) باید با استفاده از پیکرهٔ توسعه یا آموزش، تخمین زده شوند.

مدل خطی: این مدل مشابه مدل نمایی است. تفاوت آن در تابع وزن‌دهی تعریف‌شده حول s_i ها است که به‌صورت رابطه (۸) تعریف می‌شود:

$$w(k)=\begin{cases}\sum_{i|(i,j)\in A} 1-m(|k-i|) & s_k \text{ is not verb}\\ f(w(k)) & s_k \text{ is verb}\end{cases} \qquad (۸)$$

مقدار بهینهٔ پارامتر m و تابع f(.) باید با استفاده از پیکرهٔ توسعه یا آموزش تخمین زده شوند.

در این ویژگی کلمات مبدأ مرتبط با کلمهٔ مقصد در همسایگی جایگاهی یکسان با جایگاه کلمهٔ مقصد در نظر گرفته نمی‌شود. به این ترتیب این ویژگی برای جفت‌زبان‌هایی که ترتیب کلمات در جمله برای آنها متفاوت است نیز مناسب می‌باشد.

۲-۲- احتمال هم‌ترازی نقش دستوری

نقش دستوری یک کلمه مبدأ و نقش دستوری ترجمهٔ آن دارای همبستگی هستند. برای مثال در بیشتر موارد ترجمهٔ یک کلمهٔ فعل، دارای نقش دستوری فعل است و یا به‌طور معمول کلمه‌ای با نقش دستوری اسم به کلمه‌ای با همین نقش دستوری ترجمه می‌شود. منظور از نقش دستوری یک گروه‌بندی عمومی‌تر از برچسب جزء کلام است که به کلمات نقش‌های اسم، فعل، صفت، قید و ... می‌دهد. اگر $S_a=\{s_i|(i,j)\in A\}$ نشان‌دهندهٔ کلمات مبدأ هم‌ترازشده با کلمهٔ مقصد t_j باشد و SC(w) نشان‌دهنده نقش دستوری کلمه w باشد، دو ویژگی به‌صورت روابط (۹) و (۱۰) تعریف می‌شود:

$$\text{syn-align}(t_j)=P(SC(t_j)\,|\,SC(S_a))=\frac{\sum_{s_i|(i,j)\in A} N(SC(s_i),\ SC(t_j))}{\sum_{s_i|(i,j)\in A} N(SC(s_i))} \qquad (۹)$$

$$\text{backward-syn-align}(t_j)=P(SC(S_a)\,|\,SC(t_j))=\frac{\sum_{s_i|(i,j)\in A} N(SC(s_i),\ SC(t_j))}{N(SC(t_j))} \qquad (۱۰)$$

منظور از $N(sc_1,\ sc_2)$ تعداد دفعاتی است که نقش دستوری یک کلمهٔ مبدأ، sc_1 و نقش دستوری کلمهٔ ترجمه آن sc_2 باشد.

۲-۳- عبارات دستوری

به‌طور معمول ترجمهٔ یک عبارت دستوری (شامل عبارت اسمی مثل "the fight against the colonel"، عبارت صفتی مثل "weak and fragmented"، عبارت حرف اضافه مثل "during the months" و عبارت قیدی مثل " right away")، خود یک عبارت پیوسته است. بر این اساس دو ویژگی به‌صورت زیر ارائه شده است:

جایگاه کلمه در عبارات دستوری: این ویژگی به این صورت تعریف می‌شود که دو مجموعه کلمه تعریف می‌شود. مجموعه کلمات به‌جا و مجموعه کلمات نابه‌جا. کلمات مقصدی که عضو ترجمهٔ حداقل یک عبارت دستوری هستند، به روش ذکرشده در روندنمای شکل ۲ به یک یا هر دو مجموعه اضافه می‌شوند. بعد از به‌دست‌آمدن دو مجموعه طبق روندنما، مقدار ویژگی برای یک کلمه مقصد به این صورت تعیین می‌شود که برای کلماتی که عضو ترجمهٔ هیچ عبارت دستوری نیستند، مقدار "نامعلوم" و برای مابقی اگر کلمه، جزء مجموعه‌کلمات نابه‌جا بود، مقدار "بد" و در غیر این صورت مقدار "خوب" می‌گیرد.

همان‌طور که در شکل ۳ مشاهده می‌شود، عبارات "Nato**'s** operations" و "Gadhafi**'s** forces" توسط تجزیه‌گر به‌عنوان عبارات دستوری شناخته شده‌اند؛ ولی عبارات ترجمه آنها که با استفاده از هم‌ترازی‌ها به‌دست آمده‌اند، متفاوتند. عبارت "عملیات ناتو" پیوسته و عبارت "نیروهای قذافی" ناپیوسته است و کلمات {مخالفت، دولت، با، معمر} بین کلمات ترجمه عبارت واقع شده‌اند.

احتمال پیوسته‌بودن عبارات دستوری: برای محاسبهٔ ویژگی دوم مطابق رابطهٔ (۱۱) تعداد عبارات دستوری که ترجمه پیوسته دارند و کلمه مقصد عضو ترجمهٔ آنهاست بر تعداد کل عبارات دستوری که کلمه مقصد عضو ترجمه آنهاست تقسیم می‌شود. برای اینکه کلماتی که عضو ترجمهٔ هیچ عبارت دستوری نیستند، متمایز شوند برای آنها مقدار متفاوتی اختصاص می‌یابد و مقادیر عددی سایر کلمات به‌صورت گسسته استفاده می‌شوند. رابطهٔ (۱۲) نشان‌دهندهٔ این ویژگی است.

$$\text{goodness}=\frac{\#\text{continous syntactic phrases}(t_j)}{\#\text{syntactic phrases}(t_j)} \qquad (۱۱)$$

$$\text{syn.phrase.prob}(t_j)=\begin{cases}\text{no_idea} & \text{syntactic phrases}(t_j)=0\\ [\text{goodness}\times\text{discrete}_{\text{param}}] & \text{otherwise}\end{cases} \qquad (۱۲)$$

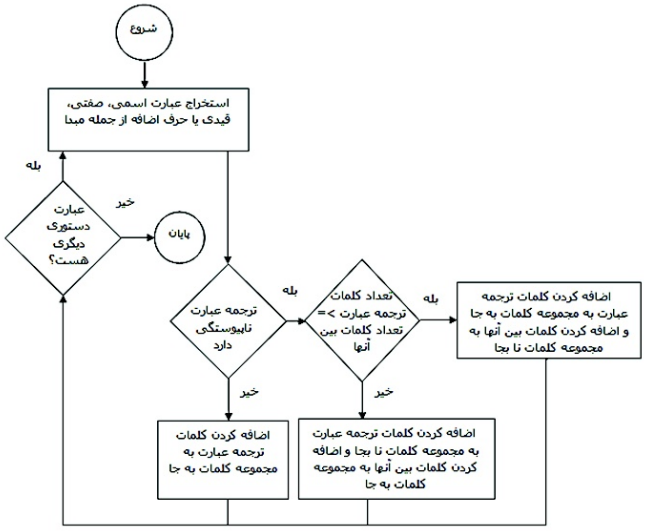
۲-۴- سازگاری فعل

یکی از خطاهای معمول ترجمه که در ترجمهٔ توسط ماشین رخ می‌دهد، سازگارنبودن فعل با جمله است. در این مقاله یک ویژگی به‌منظور بررسی سازگاری فعل جملهٔ مقصد با زمان و شخص این جمله ارائه شده است. برای برخی از زبان‌ها مانند فارسی، فعل شامل اطلاعاتی دربارهٔ زمان و شخص است. این اطلاعات ممکن است در برخی از حروف کلمهٔ فعل مانند حروف آخر آن موجود باشد. اگر این حروف را حروف بامعنی بنامیم، در این گونه موارد می‌توان هر یک از حروف بامعنی را به‌عنوان یک ویژگی در نظر گرفت؛ چون هر ترکیبی از آنها می‌تواند نشان‌دهندهٔ زمان و شخص خاصی برای فعل باشد. زمان و شخص صحیح جمله مقصد، با استفاده از جملهٔ مبدأ استخراج شده و هدف آن است که این زمان و شخص با حروف بامعنی فعل جمله مقصد مقایسه شوند. زمان و شخص فعل جملهٔ مبدأ با استفاده از برچسب جزء کلام و به‌کارگیری روش‌های سادهٔ مبتنی بر قاعده استخراج شده است. در اینجا سه حرف آخر فعل به‌عنوان حروف بامعنی در نظر گرفته شده است. شش ویژگی به‌صورت زیر استخراج می‌شوند:

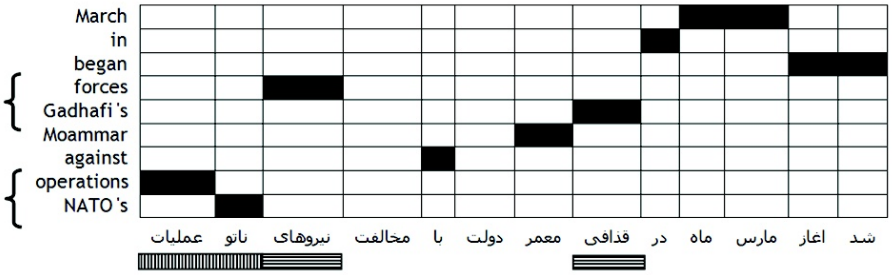
Verb Time
- Last letter + source sentence time
- Before last letter + source sentence time
- Before before last letter + source sentence time

Verb Person
- Last letter + source sentence person
- Before last letter + source sentence person
- Before before last letter + source sentence person

مزّیت این روش آن است که زمان و شخص صحیح از جملهٔ مبدأ استخراج می‌شود و با ترکیب آنها با حروف بامعنی فعل جملهٔ مقصد، درواقع صحت زمان و شخص فعل جملهٔ مقصد مورد بررسی قرار می‌گیرد. مزیت دیگر آن این است که پیاده‌سازی آن ساده است و نیازی به ابزار خاصی برای سمت مقصد ندارد؛ لذا این ویژگی برای جفت زبان‌های دیگر هم قابل استفاده است که حروف بامعنی در هر زبان متناسب با آن زبان تعیین می‌شود.



(شکل ۲): نمودار جریانی تولید مجموعه کلمات به‌جا و نابه‌جا



(شکل ۳): عبارات دستوری پیوسته و ناپیوسته

۲-۵- جایگاه نسبی کلمه در زیرجمله

جایگاه نسبی یک کلمه در زیرجملهٔ مربوط به آن (یعنی کوچک‌ترین جمله‌ای که کلمه عضو آن است) که از رابطهٔ (۱۳) به‌دست آورده می شود، در کنار برچسب جزء کلام به‌عنوان یک ویژگی در نظر گرفته شده است.

$$\text{position}(t_j)=\frac{\text{pos_in_subsent}(t_j)+1}{\text{subsent_len}+1} \qquad (۱۳)$$

مشخّص‌کردن جملات سمت مقصد با تجزیهٔ جملهٔ مبدأ و استفاده از هم‌ترازی‌های بین کلمات جملهٔ مبدأ و مقصد انجام می‌شود.

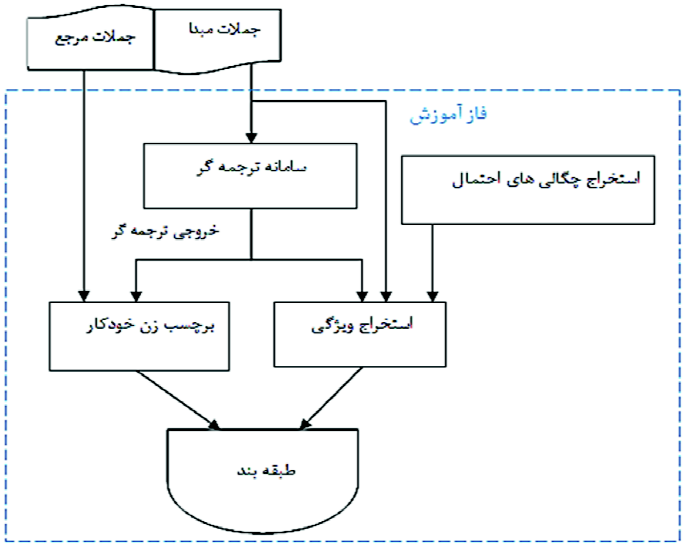
۳-نتایج تجربی

۳-۱-روش کار سامانه

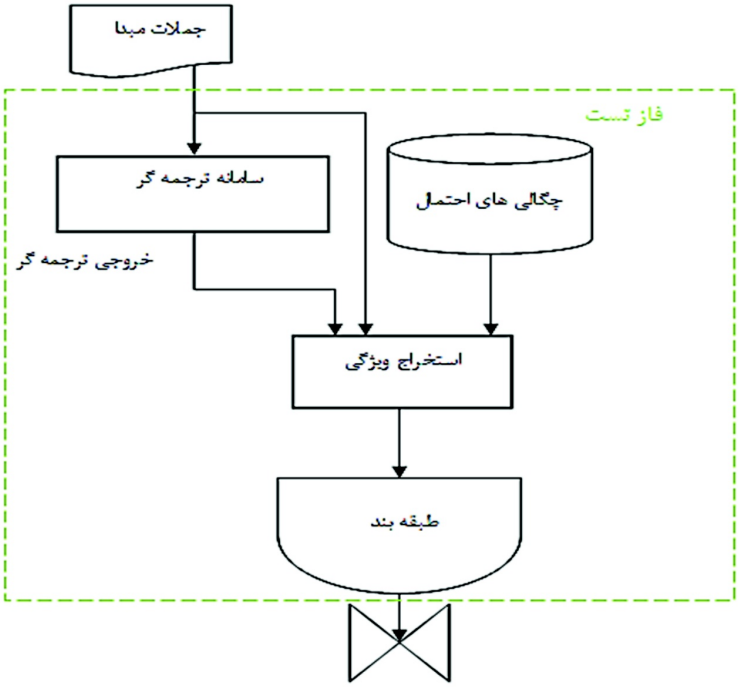
سامانهٔ تخمین اطمینان از دو مرحلهٔ آموزش و آزمون تشکیل می‌شود. در مرحلهٔ آموزش چنان‌که در شکل ۴ نشان داده شده‌است، پس از ترجمهٔ جملات مبدأ آموزش توسط سامانهٔ ترجمه‌گر، جملات خروجی ترجمه‌گر به همراه جملات مبدأ به بخش استخراج ویژگی داده می‌شوند. استخراج ویژگی با کمک منابع اطلاعاتی خارجی انجام می‌شود. برای استخراج برخی از ویژگی‌ها مانند ویژگی‌های مبتنی بر اطلاعات متقابل، لازم است از قبل چگالی‌های احتمال از روی یک پیکره به اندازهٔ کافی بزرگ به دست آمده باشند. بعد از استخراج ویژگی‌های کلمات مقصد، هر کلمهٔ مقصد به شکل یک بردار ویژگی به سامانهٔ تخمین اطمینان داده می‌شود و این سامانه با استفاده از برچسب‌های مرجع که در دسترس دارد، عمل برچسب‌زنی را آموزش می‌بینند. برچسب‌های مرجع با استفاده از خروجی ترجمه‌گر و ترجمهٔ مرجع جملات مبدأ به دست می‌آیند.

در مرحلهٔ آزمون مطابق شکل ۵ جملات مبدأ آزمون به سامانهٔ ترجمه‌گر برای ترجمه داده می‌شوند. خروجی سامانهٔ ترجمه‌گر به همراه جملات مبدأ به بخش استخراج

ویژگی داده می‌شوند. پس از استخراج بردارهای ویژگی کلمات مقصد، سامانهٔ تخمین اطمینان آموزش داده شده، به این بردارها برچسب «درست» یا «نادرست» می‌زند.



(شکل ۴): عبارات دستوری پیوسته و ناپیوسته



(شکل ۵): مرحلهٔ آزمون سامانهٔ تخمین اطمینان خروجی ترجمهٔ ماشینی

۳-۲- تهیهٔ داده‌ها

برای انجام تخمین اطمینان به داده‌های آموزشی، شامل پیکرهٔ ترجمه‌شده توسط ماشین، نیاز داریم که کلمات ترجمهٔ آن برچسب خورده باشند. برای تولید چنین پیکره‌ای سه راهکار وجود دارد (ریبود و همکاران، ۲۰۱۱):

دستی: ترجمهٔ جملات مبدأ توسط ماشین و برچسب‌زدن دستی توسط انسان

خودکار: ترجمهٔ جملات مبدأ توسط ماشین و برچسب‌زنی به‌صورت خودکار با استفاده از ترجمه‌های درست (مرجع)

نیمه‌خودکار: تولید یک پیکرهٔ برچسب‌خوردهٔ کوچک به روش یک، استخراج مدل خطای ترجمهٔ ماشین و سپس شبیه‌سازی عمل ترجمهٔ ماشین به این صورت که ترجمهٔ مرجع جملات مبدأ مورد نظر به شبیه‌ساز سامانهٔ ترجمه داده می‌شود و شبیه‌ساز با استفاده از مدل خطای به‌دست آمده در آن خطا جاسازی می‌کند.

از میان راهکارهای بالا ما به‌منظور صرفه‌جویی در زمان و هزینه برچسب‌زنی خودکار را انتخاب کردیم. روش‌های مختلفی برای این کار وجود دارد. یکی از روش‌هایی که در پژوهش‌های گذشته از جمله (بلاتز و همکاران، ۲۰۰۴) استفاده شده ، WER[5] است. در این روش، کلمه‌ای برچسب درست می‌خورد که در هم‌ترازی لونشتاین جملهٔ خروجی سامانهٔ ترجمه‌گر با جملات ترجمهٔ مرجع، در حداقل یکی از مرجع‌ها، با خودش هم‌تراز شود.

سامانهٔ ترجمه‌کنندهٔ مورد استفادهٔ ما، سامانهٔ موزز (کوهن و همکاران، ۲۰۰۳) است که یک سامانهٔ آماری مبتنی بر عبارت است. ما ترجمه‌گر موزز را با ۷۸۹۴۹۴ جفت جملهٔ موازی انگلیسی-فارسی از پیکرهٔ AFEC (جباری و همکاران، ۲۰۱۲) آموزش دادیم و سپس دو پیکرهٔ آزمون جدا را با استفاده از سامانهٔ ترجمه‌گر ترجمه کردیم. این دو پیکرهٔ آزمون، برای تهیهٔ پیکرهٔ آموزش و آزمون مربوط به تخمین اطمینان استفاده شدند. پیکرهٔ آموزش تخمین اطمینان شامل ۳۴۹۹ جمله و پیکرهٔ آزمون شامل ۱۰۳۲ جمله با چهار ترجمه مرجع است. در پیکرهٔ آموزش توزیع داده‌ها بین طبقه‌های "درست" و "نادرست" به‌ترتیب به‌صورت ۵۴ و ۴۶ درصد و در پیکره آزمون ۵۵ و ۴۵ درصد بوده است. مشخصه‌های آماری پیکره‌ها در جدول ۱ آمده است.

[5] Word Error Rate

۳-۳- معیارهای ارزیابی

برای ارزیابی سامانهٔ تخمین اطمینان جملات پیکرهٔ آزمون با استفاده از سامانهٔ مورد ارزیابی برچسب زده شده و با برچسب‌های درست (به دست آمده از روش WER) مقایسه شدند. نتایج این مقایسه با معیارهایCER[6] (اوفینگ و نای، ۲۰۰۳) ، F-measure و F-measure طبقهٔ منفی بیان شده‌اند. F-measure طبقه منفی برای کاربرد تشخیص خطای ترجمهٔ ماشینی مفید است.

(جدول ۱): مشخصه‌های آماری پیکره‌ها

		فارسی	انگلیسی
آموزش	جملات	۷۸۹۴۹۴	
	کلمات	۱۷۰۰۹۷۱۶	۱۵۴۸۹۵۸۶
	واژگان	۲۹۷۹۸۴	۳۲۷۸۳۰
توسعه	جملات	۳۴۹۷	
	کلمات	۷۹۷۳۶	۷۳۴۷۳
	کلمات خارج واژگان	۴۴	۵۳۸
آزمون۱ (آموزش‌تخمین اطمینان)	جملات	۳۴۹۹	
	کلمات	۸۰۲۹۲	۷۳۸۹۶
	کلمات خارج واژگان	۳۳	۵۹۷
آزمون۲ (آزمون‌تخمین اطمینان)	جملات	۱۰۳۲	
	کلمات	۳۰۲۴۱	۲۸۱۱۲
	کلمات خارج واژگان	۲۴	۲۴۱

۳-۴- سامانه‌های پایه

در این پژوهش از میان پژوهش‌های گذشته سه مورد انتخاب و پیاده‌سازی شد و برای پیکرهٔ انگلیسی-فارسی استفاده‌شده در این پژوهش ارزیابی شد. به این ترتیب سه سامانهٔ پایه برای مقایسهٔ نتایج پژوهش حاضر با پژوهش‌های گذشته تهیه شد. در ادامه، این سامانه‌ها مختصری شرح داده می‌شوند.

۳-۴-۱- سامانهٔ پایهٔ یک

یکی از درجه اطمینان‌های شناخته‌شده برای تخمین اطمینان احتمال پسین است که در (اوفینگ و نای، ۲۰۰۳) ارائه شد. این درجه اطمینان در تعدادی از پژوهش‌های گذشته از جمله (زیونگ و همکاران، ۲۰۱۰؛ بچ و همکاران، ۲۰۱۱) به‌عنوان سامانهٔ پایه استفاده شده است. ما احتمال پسین را با استفاده از فهرست‌های یکصد ترجمهٔ برتر محاسبه کردیم. در (اوفینگ و نای، ۲۰۰۳) دو درجهٔ

[6] Classification Error Rate

اطمینان دیگر نیز به نام‌های جمع رتبه‌ها و فراوانی مربوط ارائه شد. ما نمونه‌های آزمون را با استفاده جداگانه این سه درجه با روش انتخاب حد آستانه به‌منظور کمینه‌کردن معیار CER، طبقه‌بندی کردیم (این درجه‌ها در (اوفینگ و نای، ۲۰۰۳) نیز به‌طور جداگانه و با روش انتخاب حد آستانه استفاده شدند.).

۳-۴-۲- سامانهٔ پایهٔ دو

یکی از ویژگی‌های مبتنی بر محتوای کارا، ویژگی مبتنی بر اطلاعات متقابل بین زبانی است (ریبود و همکاران، ۲۰۰۹الف؛ ریبود و همکاران، ۲۰۰۹ب). ریبود و همکارانش علاوه‌بر این ویژگی، ویژگی مبتنی بر اطلاعات متقابل میان زبانی را نیز ارائه، اما گزارش دادند که استفاده از ویژگی مبتنی بر اطلاعات متقابل بین زبانی به‌تنهایی برای تخمین اطمینان سطح کلمه مؤثرتر است. ما برای سامانهٔ پایهٔ دو، کلمات را با استفاده از این ویژگی و روش انتخاب حد آستانه به‌منظور کمینه‌کردن CER طبقه‌بندی کردیم.

۳-۴-۳- سامانهٔ پایهٔ سه

براساس نتایج گزارش‌شده در (زیونگ و همکاران، ۲۰۱۰) ویژگی‌های لغوی ارائه‌شده پس از به‌کارگیری در یک چارچوب بیشینه آنتروپی، بهبود قابل توجهی را به دست آورند. ما نیز این ویژگی‌ها را در یک چارچوب بیشینه آنتروپی به کار بردیم.

۳-۵- ارزیابی

برای ارزیابی ویژگی‌های ارائه‌شده، ابتدا آنها را به‌صورت درجه اطمینان در نظر گرفتیم؛ به‌طوری که هر کدام از ویژگی‌ها به‌تنهایی و با روش انتخاب حد آستانه برای طبقه‌بندی نمونه‌ها مورد استفاده قرار گرفتند؛ سپس ویژگی‌های مبتنی بر ساختار با روش رگرسیون منطقی با یکدیگر ترکیب شده و با انتخاب حد آستانه برای طبقه‌بندی استفاده شدند. همچنین این ویژگی‌ها در کنار هم در چارچوب‌های درخت تصمیم و شبکهٔ عصبی RBF به کارگرفته شدند.

لازم به ذکر است، ویژگی برچسب جزء کلام که در اینجا در کنار ویژگی جایگاه کلمه در زیرجمله استفاده شده، در پژوهش‌های پیشین (ریبود و همکاران، ۲۰۱۱) به‌تنهایی استفاده شده است. از این‌رو برای اطمینان از مفیدبودن ویژگی جایگاه کلمه در زیرجمله، در جدول ۲ مقایسه‌ای بین حالتی که از ویژگی برچسب جزء کلام به‌تنهایی استفاده می شود با استفاده از آن در کنار ویژگی جایگاه کلمه در زیرجمله، آورده شده است. همان‌طورکه مشاهده می‌شود، با استفاده از ویژگی جایگاه در کنار برچسب جزء کلام معیار CER ۰/۷۶ درصد کاهش یافته است و معیار F-measure ۷/۱ درصد و F-measure طبقهٔ منفی ۱۹/۱ درصد افزایش یافته‌اند.

(جدول ۲): بررسی کارایی ویژگی جایگاه کلمه در زیرجمله در کنار برچسب جزء کلام

ویژگی‌ها	روش یادگیری	F-measure	CER	F-Measure طبقه منفی
جزء کلام	انتخاب حد آستانه	۴۷/۳	۴۲/۶۹	۱۷/۹
جزء کلام + جایگاه در زیرجمله	انتخاب حد آستانه	**۵۴/۴**	**۴۱/۹۳**	**۳۷**

در تخمین پارامترهای مدل نمایی و خطی مربوط به ویژگی مبتنی بر اطلاعات متقابل بهبودیافته، در مدل نمایی δ به‌اندازه‌ای کوچک در نظر گرفته شد که فقط کلمات همتراز و فعل وزن بگیرند. در مدل خطی نیز m به‌اندازه‌ای بزرگ در نظر گرفته شد که فقط کلمات همتراز و فعل وزن بگیرند. بنابراین برای وزن‌دهی خطی و نمایی، وزن‌ها یکسان می‌شود.

در جدول ۳ نتایج مربوط به طبقه‌بندی نمونه‌های آزمون با این ویژگی با استفاده از مقادیر بهینهٔ پارامترها آمده است. نتایج مربوط به استفاده از سایر ویژگی‌ها، به‌تنهایی نیز در این جدول قابل مشاهده است. همان‌گونه که مشاهده می‌شود، همه ویژگی‌های ارائه‌شده، در معیار CER نسبت به سامانهٔ پایهٔ یک برتری دارند. از آنجا که در روش طبقه‌بندی استفاده‌شده حد آستانه با هدف کمینه‌کردن CER تعیین شده است، می‌توان نتیجه گرفت که ویژگی‌های ارائه‌شده قابلیت یادگیری سامانه را در راستای معیار هدف بیشتر می‌کنند. در ویژگی‌های سازگاری فعل زمان شامل "ماضی"، "مضارع" و "آینده" و شخص شامل "من"، "تو"، "او"، "ما" و "ایشان" در نظر گرفته شد. این ویژگی‌ها نتایج مطلوبی نداده‌اند. یکی از دلایل آن می‌تواند این باشد که قواعد ساده‌ای که برای استخراج زمان و شخص فعل استفاده شده، دقت کافی را نداشته‌اند. ویژگی احتمال هم‌ترازی طبقه کلمات پسرو علاوه‌بر برتری نسبت به سامانه‌های پایهٔ یک در همهٔ معیارها،

نسبت به سامانهٔ پایه سه بهبود ۶/۱۵ درصدی در معیار F-measure طبقهٔ منفی ایجاد کرده است. ویژگی جایگاه کلمه عبارت دستوری، علاوه‌بر برتری نسبت به سامانه‌های پایهٔ یک و سه در همه معیارها، نسبت به سامانهٔ پایهٔ دو نیز در معیارهای CER و F-measure به‌ترتیب برتری ۱/۴۷ و هشت درصدی داشته است. دو ویژگی احتمال دستوری‌بودن عبارت و ویژگی مبتنی بر اطلاعات متقابل بهبودیافته نسبت به همه سامانه‌های پایه در تمامی معیارها برتری داشته‌اند. ویژگی مبتنی بر اطلاعات متقابل بهبودیافته نسبت به بهترین سامانهٔ پایه (۲) برتری ۹/۶۳ درصدی در CER، ۸/۵ درصدی در F-measure و ۵/۱ درصدی در F-measure طبقه منفی داشته است. عملکرد قوی این ویژگی نسبت به ویژگی مبتنی بر اطلاعات متقابل بین زبانی (ریبود و همکاران، ۲۰۰۹ب) مورد انتظار بود، چون جفت زبان فارسی-انگلیسی دارای ترتیب کلمات متفاوتی هستند و وزن‌دهی ارائه‌شده در این مقاله برای آنها مناسب‌تر است.

در جدول ۴ ترکیب ویژگی‌های مبتنی بر ساختار با روش‌های مختلف مورد ارزیابی قرار گرفته است. در آزمایش‌های انجام‌شده مشخص شد ویژگی‌های زمان و شخص فعل در ترکیب با سایر ویژگی‌ها بهبود قابل توجهی ایجاد نمی‌کنند؛ از این‌رو در جدول ۴ نتایج مربوط به ترکیب سایر ویژگی‌های مبتنی بر ساختار آمده است. به‌دلیل مستقل‌نبودن ویژگی‌ها از یکدیگر، در ترکیب ویژگی‌ها از روش‌های تمایزی استفاده شده است. نتایج نشان می‌دهد روش‌های درخت تصمیم و شبکهٔ عصبی RBF بهترین کارایی را حاصل می‌کنند. استفاده از ترکیب ویژگی‌های مبتنی بر ساختار ارائه‌شده در این مقاله، در مقایسه با بهترین سامانهٔ پایه به‌ترتیب بهبود ۴/۵۹، ۴/۱ و ۲ درصدی در معیارهای CER، F-measure و F-measure طبقه منفی ایجاد کرده است. ترکیب ویژگی‌های مبتنی بر ساختار با ویژگی مبتنی بر محتوای ارائه‌شده کارایی بهتری نسبت به استفاده از ویژگی مبتنی بر محتوا به‌تنهایی، ایجاد نمی‌کند.

(جدول ۳): نتایج مربوط به استفاده از درجه‌های اطمینان در مقایسه با سامانه‌های پایه

	ویژگی‌ها	روش یادگیری	F-measure	CER	F-Measure طبقه منفی
سامانه پایه۱	احتمال پسین	انتخاب حد آستانه	۵۰/۱	۴۶/۹۱	۳۳/۴
	فرکانس مربوط	انتخاب حد آستانه	۵۱/۶	۴۸/۵۶	۴۸/۲
	جمع رتبه	انتخاب حد آستانه	۵۱/۶	۴۸/۵۰	۴۷/۸
سامانه پایه۲	اطلاعات متقابل بین‌زبانی	انتخاب حد آستانه	۶۱/۸	**۳۸/۰۳**	**۵۶/۲**
سامانه پایه۳	لغوی	بیشینه انتروپی	۶۰/۸۵	۳۸/۵۲	۴۹/۵۵
ویژگی‌های مبتنی بر ساختار ارائه‌شده	احتمال هم‌ترازی دستوری عقب‌رو	انتخاب حد آستانه	۶۱/۱	۳۹/۱۹	۵۵/۷
	احتمال هم‌ترازی دستوری	انتخاب حد آستانه	۴۳/۳	۴۱/۴۴	۰
	جزء کلام + جایگاه در زیرجمله	انتخاب حد آستانه	۵۴/۴	۴۱/۹۳	۳۷
	زمان	انتخاب حد آستانه	۴۵/۷	۴۲/۲۷	۱۴
	شخص	انتخاب حد آستانه	۴۵/۵	۴۲/۷۸	۱۳/۶
	احتمال پیوسته‌بودن عبارت دستوری	انتخاب حد آستانه	۶۲/۴	۳۷/۶۶	۵۹
	جایگاه در عبارت دستوری	انتخاب حد آستانه	۶۲/۶	۳۶/۵۶	۵۳/۸
ویژگی‌های مبتنی بر محتوای ارائه‌شده	اطلاعات متقابل بهبود یافته	انتخاب حد آستانه	**۷۰/۳**	**۲۸/۴۰**	**۶۱/۳**

(جدول ۴): نتایج مربوط به ترکیب ویژگی‌های مبتنی بر ساختار

ویژگی‌ها	روش یادگیری	F-measure	CER	F-Measure طبقه منفی
احتمال هم‌ترازی دستوری+ عقب‌رو +	رگرسیون منطقی	۶۵	۳۴/۲۷	۵۶/۹
جزء کلام + جایگاه در زیرجمله +	درخت تصمیم	۶۵/۹	۳۳/۴۴	۵۸/۲
عبارات دستوری	RBF	۶۵/۷	۳۳/۶۳	۵۸/۲

۴- نتیجه‌گیری

در این مقاله به‌منظور داشتن سامانهٔ تخمین اطمینانی مستقل از سامانط ترجمه‌گر، تمرکز بر روی ویژگی‌های مستقل بود و پنج گروه ویژگی جدید در قالب ویژگی‌های مبتنی بر محتوا و مبتنی بر ساختار ارائه شد. برای مقابله با چالش نقص ساختاری جملات مقصد در ویژگی‌های مبتنی بر ساختار، سازگاری ساختاری کلمات مقصد با جملات مبدأ مورد بررسی قرار گرفت. درنتیجه این ویژگی‌ها تنها نیاز به تجزیهٔ نحوی جملات مبدأ دارند. همچنین با ایجاد تغییرات منطقی در یک ویژگی مبتنی بر محتوای کارامد، این ویژگی برای جفت زبان‌های با ترتیب کلمات ناهمگون مثل انگلیسی-فارسی بهبود قابل توجهی یافت. نتایج نشان می‌دهد، ویژگی متنی بر محتوا نسبت به بهترین سامانهٔ پایه (مبتنی بر اطلاعات متقابل بین زبانی) ۹/۶۳ درصد در CER، ۸/۵ درصد در F-measure و ۵/۱ درصد در F-measure طبقهٔ منفی برتری داشته است. ترکیب ویژگی‌های مبتنی بر ساختار ارائه‌شده نیز در مقایسه با بهترین سامانهٔ پایه، بهبود ایجاد کرده است که میزان آن ۴/۵۹ درصد در CER، ۴/۱ درصد در F-measure و دو درصد در F-measure طبقه منفی است.

۵- مراجع

Bach, N., F. Huang and Y. Al-Onaizan (2011). Goodness: A method for measuring machine translation confidence. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, Association for Computational Linguistics.

Blatz, J., E. Fitzgerald, G. Foster, S. Gandrabur, C. Goutte, A. Kulesza, A. Sanchis and N. Ueffing (2004). Confidence estimation for machine translation. Proceedings of the 20th international conference on Computational Linguistics, Association for Computational Linguistics.

Buck, C. (2012). Black box features for the WMT 2012 quality estimation shared task. Proceedings of the Seventh Workshop on Statistical Machine Translation, Association for Computational Linguistics.

FattanehJabbari, S. and S. M. M. Ziabary (2012). Developing an Open-domain English-Farsi Translation System Using AFEC: Amirkabir Bilin-gual Farsi-English Corpus. The Fourth Wor-kshop on Computational Approaches to Arabic Script-based Languages.

Guo, G., C. Huang, H. Jiang and R.-H. Wang (2004). A comparative study on various confidence measures in large vocabulary speech recognition. Chinese Spoken Language Processing, 2004 International Symposium on, IEEE.

Hildebrand, A. S. and S. Vogel (2008). Combination of machine translation systems via hypothesis selection from combined n-best lists. MT at work: Proceedings of the Eighth Conference of the Association for Machine Translation in the Americas, Citeseer.

Koehn, P., F. J. Och and D. Marcu (2003). Statistical phrase-based translation. Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1, Association for Computational Linguistics.

Luong, N.-Q., L. Besacier, C. d. G. LIG and B. Lecouteux (2014). "Lig system for word level qe task at wmt14." ACL 2014: 335.

Luong, N. Q., B. Lecouteux and L. Besacier (2013). LIG system for WMT13 QE task: Investigating the usefulness of features in word confidence estimation for MT. Proceedings of the Eighth Workshop on Statistical Machine Translation, Citeseer.

Quirk, C. (2004). Training a Sentence-Level Machine Translation Confidence Measure. LREC, Citeseer.

Raybaud, S., D. Langlois and K. Smaïli (2009). Efficient combination of confidence measures for machine translation. 10th Annual Conference of the International Speech Communication Association-INTERSPEECH 2009.

Raybaud, S., D. Langlois and K. Smaïli (2011). "“This sentence is wrong.” Detecting errors in machine-translated sentences." Machine translation **25**(1): 1-34.

Raybaud, S., C. Lavecchia, D. Langlois and K. Smaïli (2009). "New confidence measures for statistical machine translation." arXiv preprint arXiv:0-902.1033.

Raybaud, S., C. Lavecchia, D. Langlois and K. Smaili (2009). Word-and sentence-level confidence measures for machine translation. 13th Annual Meeting of the European Association for Machine Translation-EAMT 09.

Specia, L., C. Saunders, M. Turchi, Z. Wang and J. Shawe-Taylor (2009). "Improving the confidence of machine translation quality estimates." Proceedings of the Twelfth Machine Translation Summit (MT Summit XII): 136-143.

Specia, L., M. Turchi, N. Cancedda, M. Dymetman and N. Cristianini (2009). Estimating the sentence-level quality of machine translation systems. 13th Conference of the European Association for Machine Translation.

Ueffing, N., K. Macherey and H. Ney (2003). Confidence measures for statistical machine tran-slation. In Proc. MT Summit IX, Citeseer.

Ueffing, N. and H. Ney (2005). Application of word-level confidence measures in interactive statistical machine translation. Proc. EAMT.

Xiong, D., M. Zhang and H. Li (2010). Error detection for statistical machine translation using linguistic features. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics.

مرضیه صالحی مدرک کارشناسی خود را در رشتهٔ مهندسی کامپیوتر از دانشگاه علم و صنعت ایران در سال ۱۳۸۹ و مدرک کارشناسی ارشد خود را در رشتهٔ هوش مصنوعی در سال ۱۳۹۳ از دانشگاه الزهرا (س) دریافت کرده است.
نشانی رایانامهٔ ایشان عبارت است از:
marzieh.salehi.sh@gmail.com

شهرام خدیوی مدارک کارشناسی و کارشناسی ارشد خود را در رشتهٔ مهندسی کامپیوتر از دانشگاه صنعتی امیرکبیر به‌ترتیب در سال‌های ۱۳۷۵ و ۱۳۷۸ دریافت کرده و مدرک دکترای خود را در سال ۱۳۸۷ از دانشگاه آخن RWTH در رشتهٔ علوم کامپیوتر دریافت کرده‌اند. ایشان در حال حاضر عضو هیأت علمی و استادیار دانشگاه صنعتی امیرکبیر هستند. زمینه‌های پژوهشی مورد علاقهٔ ایشان پردازش زبان طبیعی، ترجمهٔ ماشینی آماری و یادگیری ماشین است.
نشانی رایانامهٔ ایشان عبارت است از:
khadivi@aut.ac.ir

نوشین ریاحی مدرک کارشناسی خود را در سال ۱۳۶۷ از دانشگاه صنعتی اصفهان و مدارک کارشناسی ارشد و دکترای خود را از دانشگاه صنعتی شریف به‌ترتیب در سال‌های ۱۳۷۰ و ۱۳۷۹ در رشتهٔ مهندسی برق-الکترونیک دریافت کرده‌اند. ایشان در حال حاضر عضو هیأت علمی و استادیار دانشگاه الزهرا (س) هستند. زمینه‌های پژوهشی مورد علاقهٔ ایشان طراحی و ساخت ICهای باند رادیویی، بازشناسی و پردازش گفتار و ترجمهٔ ماشینی است.
نشانی رایانامهٔ ایشان عبارت است از:
nriahi@alzahra.ac.ir