دانشگاه الزهرا (س)

دانشکده فنی و مهندسی

پایان نامه

جهت اخذ درجه کارشناسی ارشد

رشته مهندسی کامپیوتر- گرایش هوش مصنوعی

عنوان

تخمین اطمینان خروجی ترجمه ماشینی

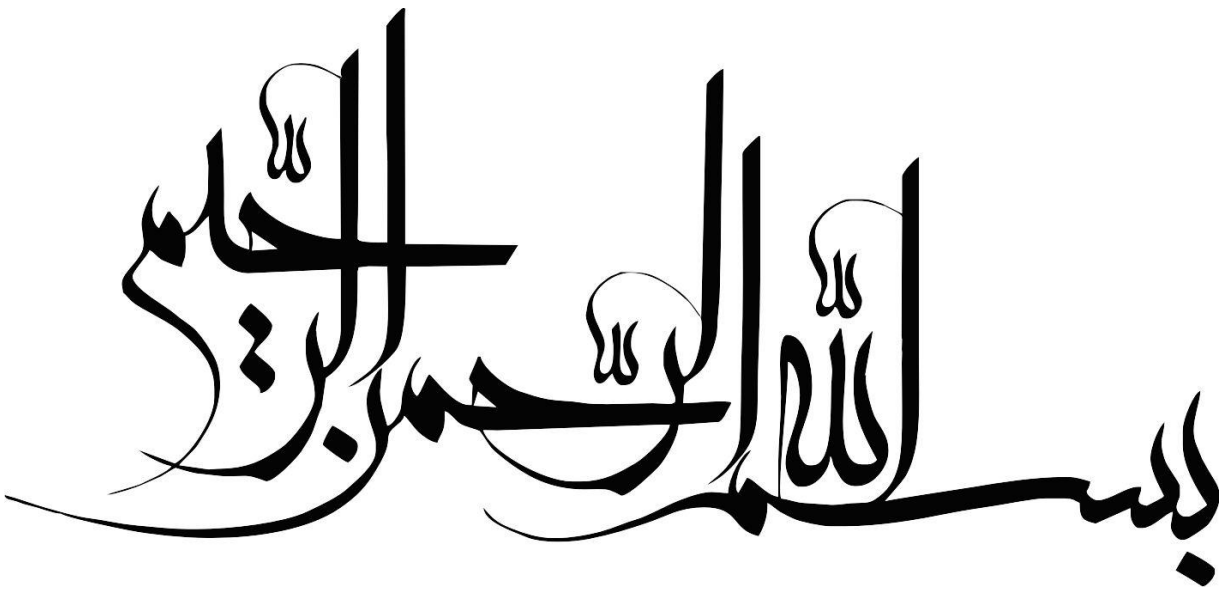
استاد راهنما

دکتر نوشین ریاحی

دانشجو

مرضیه صالحی

اسفند ۱۳۹۲


بسم الله الرحمن الرحيم

دانشگاه الزهرا (س)

دانشکده فنی و مهندسی

پایان نامه

جهت اخذ درجه کارشناسی ارشد

رشته مهندسی کامپیوتر- گرایش هوش مصنوعی

عنوان

تخمین اطمینان خروجی ترجمه ماشینی

استاد راهنما

دکتر نوشین ریاحی

استاد مشاور

دکتر شهرام خدیوی

دانشجو

مرضیه صالحی

اسفند ۱۳۹۲

کلیه دستاوردهای این تحقیق متعلق به دانشگاه الزهرا (س)

می‌باشد.

سپاس و ستایش خداوند را که مرا در این راه یاری کرد.

با تشکر از

اساتید بزرگوارم سرکار خانم دکتر ریاحی و جناب آقای دکتر خدیوی که مرا در انجام این پژوهش راهنمایی کردند،

همسر فداکارم که در این راه همراه من بود،

و پدر و مادر عزیزم که دلسوزانه پشتیبانم بودند.

چکیده

به دلیل ابهام ذاتی موجود در زبان طبیعی، تقریبا همه فناوری‌های پردازش زبان طبیعی ناکاملند. با این حال با تخمینی از کیفیت خروجی، کاربران می‌توانند به طور مناسبی با ناکامل بودن آنها برخورد کنند. زمینه تحقیقاتی که به این مسئله می‌پردازد تخمین اطمینان نامیده می‌شود. هدف اصلی تخمین اطمینان کاربردی‌تر کردن فناوری‌های ناکامل است. در حوزه ترجمه ماشینی نیز با وجود پیشرفت‌های به دست آمده در سال‌های اخیر، این فناوری هنوز قادر به ترجمه دقیق متون نیست به طوری که گاهی ممکن است ترجمه متنی توسط ماشین و ویرایش خروجی توسط انسان، زمان بیشتری نسبت به ترجمه مستقیم توسط انسان بگیرد. در چنین حالتی، داشتن تخمینی از درستی خروجی ترجمه ماشینی برای ویرایشگران مفید است تا تلاش خود را به جملات نادرستی معطوف کنند که نیاز به تغییرات پرهزینه ندارند. علاوه بر پس‌ویرایش، تخمین اطمینان خروجی ترجمه ماشینی برای کاربردهایی که هدف آنها ارتقا کیفیت ترجمه ماشینی است، ازقبیل ترکیب چند سامانه ترجمه‌گر، بازتولید خروجی و مرتب سازی دوباره لیست چند ترجمه برتر، مفید می‌باشد. تخمین اطمینان یا سنجش کیفیت خروجی ترجمه ماشینی یکی از موضوعات چالشی در زمینه ترجمه ماشینی محسوب می‌شود. همچنین برای جفت زبان انگلیسی-فارسی پژوهش‌های چندانی در زمینه تخمین اطمینان صورت نگرفته است. در این پژوهش مجموعه‌ای از ویژگی‌های مبتنی بر ساختار و مبتنی بر محتوای مستقل از سامانه ترجمه‌گر ارائه شده و کارایی چند روش یادگیری برای ترکیب این ویژگی‌ها بررسی شده است. مزیت ویژگی‌های ساختاری ارائه شده این است که برخلاف اکثر ویژگی‌های ارائه شده در گذشته، برای بررسی صحت ساختاری کلمه مقصد از جمله مبدا استفاده می‌کنند. همچنین برای نخستین بار از بردار زمینه برای تخمین اطمینان استفاده شده است و چالش متفاوت بودن فضای بردارهای مبدا و مقصد، با استفاده از روشی آماری حل شده است. نتایج به دست آمده از آزمون ویژگی‌های پیشنهادی در قالب جداولی ارائه گردیده است.

کلمات کلیدی: تخمین اطمینان، ترجمه ماشینی، بردار زمینه، اطلاعات متقابل، ویژگی‌های مبتنی بر ساختار، ویژگی‌های مبتنی بر محتوا.

فهرست مطالب

فهرست شکل ها

فهرست جدول ها

فصل اول:

مقدمه

۱-۱ ترجمه ماشینی آماری

ترجمه ماشینی آماری[1] را می‌توان به صورت یادگیری عمل ترجمه کردن از روی داده‌های آموزشی توسط ماشین، تعریف کرد. این کار به صورت طراحی مدل و یادگیری پارامترهای مدل از روی داده‌ها انجام می‌شود. در ترجمه ماشینی آماری، ترجمه به صورت یک فرآیند تصمیم گیری مدل می‌شود. با داشتن یک جمله مبدا $f_1^J = f_1 \dots f_j \dots f_J$، جمله مقصد $e_1^I = e_1 \dots e_i \dots e_I$ با بیشینه کردن احتمال پسین به دست می‌آید:

$$\hat{e}_1^I = \underset{I,e_1^I}{\operatorname{argmax}}\{\Pr(e_1^I|f_1^J)\} = \underset{I,e_1^I}{argmax}\{Pr(f_1^J|e_1^I).Pr(e_1^I)\} \qquad (۱-۱)$$

بنابراین به دو مدل نیاز است. مدل ترجمه $Pr(f_1^I|e_1^I)$ و مدل زبانی $Pr(e_1^I)$ که آنها را می‌توان به طور مستقل مدل کرد. مدل زبانی جمله مقصد خوش ترتیبی جمله مقصد را در بر می‌گیرد. مدل‌های ترجمه را می‌توان به صورت زیر تقسیم بندی کرد [1]:

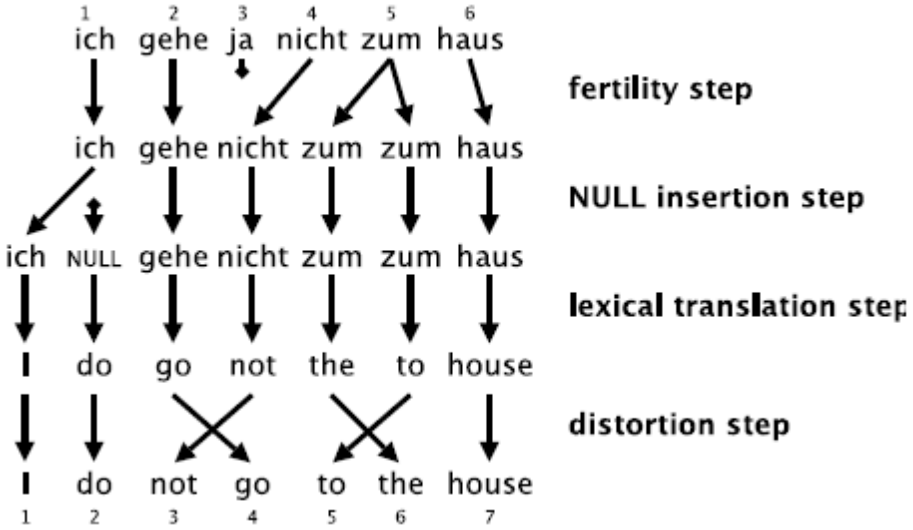
۱. مدل‌های مبتنی بر کلمه
۲. مدل‌های مبتنی بر عبارت
۳. مدل‌های مبتنی بر درخت

امروزه بیشتر سامانه‌های ترجمه‌گر آماری بر مبنای عبارات دوزبانه هستند [2, 3, 4, 5, 6, 7, 8]. از اینرو این مدل با جزئیات بیشتری بیان می‌شود.

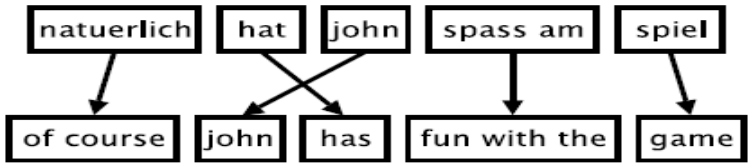
الف) مدل‌های مبتنی بر کلمه: مدل‌هایی هستند که کوچکترین واحدی که مدل‌ها بر اساس آنها ساخته می‌شود کلمه است. مثلا مدل‌هایی که شامل احتمال‌هایی از این قبیلند: احتمال اینکه یک کلمه خاص، ترجمه کلمه خاص دیگری باشد. در شکل ۱- ۱ شمایی از این مدل آمده است. این شکل مربوط به مدل IBM-3 است.

ب) مدل‌های مبتنی بر عبارت: در این مدل‌ها احتمالات بر روی واحد بزرگتری به نام عبارت تعریف می‌شوند. عبارت به معنی دنباله‌ای از کلمات است و لزوما به معنی عبارتی که از لحاظ زبان معتبر باشد نیست. عبارت‌های دوزبانه از پیکره دوزبانه آموزش که همترازی کلمات روی آن انجام شده‌است به دست می‌آیند. در شکل ۱- ۲ شمایی از مدل مبتنی بر عبارت دیده می‌شود.

[1] Statistical machine translation



شکل ۱- ۱: مدل مبتنی بر کلمه [1]



شکل ۱- ۲: مدل مبتنی بر عبارت [1]

احتمال پسین $P(f_1^I|e_1^I)$ مستقیما با استفاده از ترکیب لگاریتم-خطی وزن دار مدل زبانی، مدل ترجمه عبارت و مدل لغوی مبتنی بر کلمه به دست می‌آید.

در روش مبتنی بر عبارت، سه مرحله ضروری است: قطعه‌بندی جمله ورودی، ترجمه هر عبارت به عبارت معادل زبان مقصد و تعیین ترتیب درست عبارات ترجمه شده در جمله مقصد.

ابتدا جمله ورودی f به دنباله‌ای از I عبارت $\bar{f}_1^{\mathrm{I}}$ تقسیم می‌شود. سپس هر عبارت $\bar{f}_{\mathrm{i}}$ به عبارت معادل $\bar{e}_{\mathrm{i}}$ ترجمه می‌شود. در نهایت ممکن است عبارات ساخته شده جابجا شوند. برای مثال در یکی از سامانه‌های مبتنی بر عبارت شناخته شده به نام موزز [3]، از ۴ مدل برای ترجمه استفاده می‌شود:

- جدول عبارات $\phi(\bar{f}_i|\bar{e}_i)$
- مدل زبانی $P_{LM}(e)$

- مدل جابجایی[1]
- جریمه کلمه[2]

هر یک از موارد فوق، بخشی از کیفیت ترجمه را کنترل می‌کند. مدل جابجایی، ترتیب عبارات در جمله را کنترل می‌کند و با هر جابجایی هزینه‌ای را به سامانه تحمیل می‌نماید.

جریمه کلمه، وظیفه کنترل تعداد کلمات در جمله را بر عهده دارد و اجازه نمی‌دهد که جملات با طول بسیار بزرگ و یا بسیار کوچک تولید شوند. به هر یک از اجزاء فوق وزنی نسبت داده می شود و احتمال ترجمه دو عبارت به شکل زیر محاسبه می‌گردد:

$$P(e|f) = \phi(f|e)^{weight_{\phi}} . LM^{weight_{LM}} . D(e,f)^{weight_d} . W(e)^{weight_w} \quad (۲-۱)$$

بطوریکه $P(f|e)$ در پیاده‌سازی موزز به شکل زیر تجزیه می‌شود:

$$P(e \mid f) = P(f \mid e) P_{LM}(e) \omega^{length(e)} \quad (۳-۱)$$

بر اساس ترجمه مبتنی بر عبارت ابتدا جمله ورودی f به I عبارت $\bar{f}_1^I$ افراز می‌شود سپس هر عبارت $\bar{f}_i$ به عبارت معادل $\bar{e}_i$ از زبان مقصد ترجمه می‌شود. در نهایت عبارات در جمله معادل جابجا خواهند شد. احتمال ترجمه عبارت با تابع $\emptyset(\bar{f}_i|\bar{e}_i)$ بیان می‌شود. جابجایی عبارات خروجی در حالت پیش فرض با استفاده از مدل ساده $d(start_i, end_{i-1})$ محاسبه می‌شود. $start_i$ ابتدای عبارتی است که به i امین عبارت از زبان مقصد ترجمه شده و end_{i-1} انتهای عبارتی است که به *i-1* امین عبارت از زبان مقصد ترجمه شده‌است. ساده ترین روش تعریف مدل جابجایی به صورت $d(start_i, end_{i-1}) = \alpha^{|start_i - end_{i-1} - 1|}$، با انتخاب مقدار مناسب برای α است. در نهایت بهترین ترجمه اینگونه محاسبه می‌شود:

$$\begin{aligned} e_{best} &= \arg\max_e P(e \mid f) \\ &= \arg\max_e P(f \mid e) P_{LM}(e) \omega^{length(e)} \end{aligned} \quad (۴-۱)$$

بطوریکه:

$$P(\bar{f}_1^I \mid \bar{e}_1^I) = \Phi(\bar{f}_i \mid \bar{e}_i) d(start_i, end_{i-1}) \quad (۵-۱)$$

[1] Distortion Model

[2] Word Penalty

برای یافتن بهترین دنباله‌ای که بیشترین احتمال ترجمه را دارا باشد، لازم است فضای حالات ممکن جستجو گردد. به دلیل گستردگی فرضیات خروجی، جستجوی این فضا از مرتبه نمایی بوده و حل مسئله یافتن جمله با حداکثر احتمال، یک مسئله NP-Complete است. بنابراین برای حل بهینه این مسئله از الگوریتم‌های جستجوی بهینه مانند جستجوی شعاعی[1] یا A* استفاده می‌شود.

ارتقاء کیفیت ترجمه ابتدا با جدول عبارات دقیق‌تر و سپس تنظیم مناسب‌تر پارامترهای رمزگشایی حاصل می‌شود. برای این منظور ابتدا مدل بهتری تخمین زده می‌شود. مثلا مدل زبانی مرتبه بالاتر استفاده می‌گردد. مدل زبانی از دو جنبه موثر است. اول در محاسبه هزینه آتی[2] برای امتیاز دهی عبارات، که با بهبود آن عبارات بهتری تولید خواهد شد. از طرف دیگر برای اعتبارسنجی رشته خروجی تولید شده و برای کنترل کیفیت ترجمه نهایی استفاده می‌گردد و بنابراین در دو مرحله می‌تواند به بهبود سامانه منجر شود. اما نکته این است که با ارتقاء و بهبود مدل به سمت مدل بالاتر که لاجرم پیچیدگی بیشتری را نیز به همراه خواهد داشت، ضروری است که فضای جستجو نیز متناسب با آن افزایش داده شود.

مدل‌های مبتنی بر درخت: در مدل‌های مبتنی بر درخت ساختار دستور زبانی جملات نیز مورد توجه قرار می‌گیرد و شرط اعتبار عبارت‌های مدل قبل موجود بودن تمام کلمات عبارت در یک زیر درخت دستوری از جمله است. شکل ۱- ۳ شمایی از این مدل را نمایش می‌دهد.

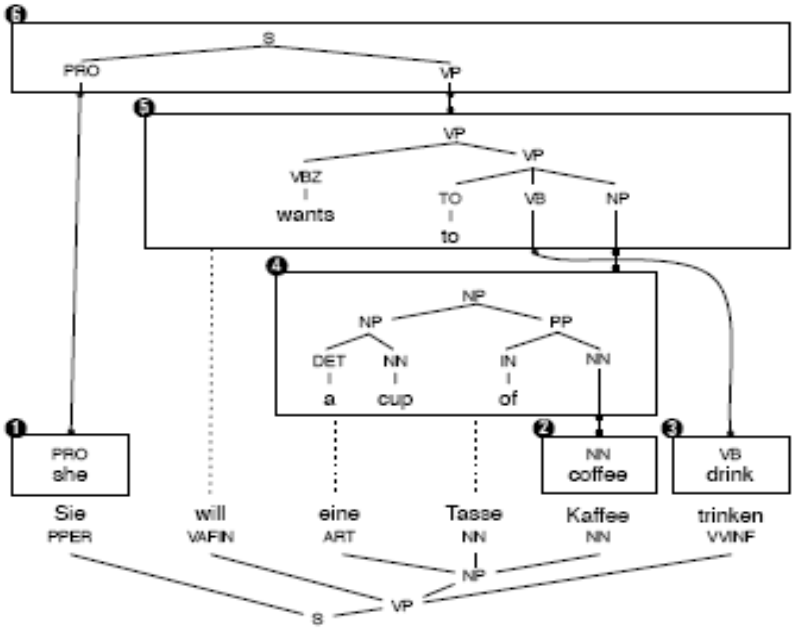
۱-۲ تخمین اطمینان

تخمین اطمینان (CE) یک رویکرد یادگیری ماشین برای محاسبه درجه درست بودن خروجی (درجه اطمینان) در یک کاربرد مبتنی بر فناوری‌های ناکامل، مانند ترجمه ماشینی است. اگر $x \in X$ ورودی ترجمه‌گر و $y \in Y$ خروجی مربوط به آن در نظر گرفته شود، درجه اطمینان تابعی به صورت $c : X \times Y \times K \rightarrow I$ است، به طوری که X دامنه ورودی ها، Y دامنه خروجی ها، K دامنه دانش اضافه مورد نیاز و I مجموعه مقادیری که نماینده درستی خروجی هستند، می‌باشند. I می‌تواند بازه‌ای از اعداد حقیقی یا مجموعه‌ای از سطوح مختلف مانند {«افزوده شده»، «جابه جا شده»، «جانشین شده»، «خوب»} [9] و یا مجموعه دوتایی {«درست»، «نادرست»} [10, 11, 12, 13, 14] باشد. یک حالت خاص از تابع c به صورت $P(C = 1|x, y, k)$ تعریف می‌شود که $C \in \{0,1\}$. $C = 1$ نشان دهنده درستی y است. تخمین اطمینان به طور گسترده در حوزه تشخیص

[1] Beam search

[2] Future cost

گفتار استفاده شده‌است، اما در حوزه‌های دیگر پردازش زبان طبیعی پیشینه کوتاه‌تری دارد. تخمین اطمینان را می‌توان به صورت لایه‌ای در نظر گرفت که بر روی سامانه پایه قرار می‌گیرد، برای مثال در تشخیص گفتار خروجی‌های سامانه تشخیص گفتار به این لایه داده می‌شود و در صورتی که درجه اطمینان پایینی برای آن تخمین زده شود، از کاربر تقاضا می‌شود جمله را دوباره تکرار کند. تخمین اطمینان می‌تواند در سطوح مختلفی از جمله سطح کلمه، سطح زیرجمله [15]، سطح جمله [16] و سطح سند (به ندرت) انجام شود. هدف تخمین اطمینان در هریک از این سطوح زدن برچسب به واحد مربوط به آن سطح است.



شکل ۱- ۳: مدل مبتنی بر درخت [1]

اگرچه تخمین اطمینان و ارزیابی خودکار ترجمه ماشینی با استفاده از معیارهایی از قبیل BLEU [17] و NIST [18]، هر دو کیفیت خروجی ترجمه ماشینی را تعیین می‌کنند، ولی با هم متفاوتند. مهم ترین تفاوت بین آنها ناشی از این واقعیت است که تخمین اطمینان بدون مقایسه خروجی ترجمه ماشینی با ترجمه‌های درست و یا پرسیدن نظر یک مترجم انسانی در مورد خروجی انجام می‌شود.

۱-۲-۱ تخمین اطمینان ترجمه ماشینی در سطح کلمه

هدف تخمین اطمینان خروجی ترجمه ماشینی در سطح کلمه نسبت دادن احتمال درستی یا امتیاز درستی (تخمین اطمینان قوی) و یا برچسب‌های گسسته (تخمین اطمینان ضعیف) به کلمات خروجی ترجمه ماشینی است که بدون استفاده از ترجمه مرجع انجام می‌شود. بر اساس نوع تخمین اطمینانی که تولید می‌شود (ضعیف یا قوی)، الگوریتم‌های یادگیری ماشین متفاوتی برای این کار استفاده می‌شوند. برای تولید تخمین اطمینان ضعیف، الگوریتم‌های طبقه بندی یادگیری ماشین برای نسبت دادن برچسب دوتایی یا چندتایی به هر کلمه، مورد استفاده قرار می‌گیرند. در حالت قوی، تکنیک‌های رگرسیون به کار می‌روند. در این بخش جزئیاتی از تخمین اطمینان سطح کلمه با استفاده از یک روش یادگیری نظارتی بیان می‌شود.

برای یک کلمه در جایگاه j از یک جمله خروجی ترجمه ماشینی، برچسب تخمین اطمینان آن به صورت زیر نشان داده می‌شود که t جمله مقصد و s جمله مبدا می‌باشد:

$$C_{j,t,s} \in \{0,1\} \qquad (۶-۱)$$

برای تولید تخمین اطمینان تابع توزیع احتمال درستی یک کلمه در حالی که جمله مبدا و مقصد داده شده‌اند، یادگیری می‌شود. از نظر ریاضی، با داشتن جمله مبدا و مقصد، هدف یادگیری، تخمین احتمال زیر برای هر کلمه است:

$$p(C_{j,t,s} = 1|j,t,s) \qquad (۷-۱)$$

اما چون s و t ممکن است هر جفت جمله‌ای باشند، احتمال بالا به صورت مستقیم به دست نمی‌آید. به همین دلیل هر کلمه به صورت یک بردار از ویژگی‌های عددی نشان داده می‌شود. چنین بردار ویژگی با استفاده از یک تابع x به دست می‌آید. تابع x اطلاعات استخراج شده از جملات مبدا و مقصد و خروجی ترجمه ماشینی را به بردار ویژگی نظیر می‌کند:

$$x\colon (j,t,s) \in j \times l_s \times l_t \rightarrow x(j,t,s) \in R^{d_w} \qquad (۸-۱)$$

در واقع اطلاعات استخراج شده از جفت جمله مبدا و مقصد و خروجی ترجمه ماشینی، برای استخراج ویژگی‌هایی که جنبه‌های مختلف ترجمه را در بر می‌گیرند به کار می‌رود. با نمایش کلمات به این شیوه، تابع توزیع احتمال یادگیری شده به صورت زیر در می‌آید:

$$p\big(C_{j,t,s}; j,t,s\big) = P(C_{j,t,s}|x(j,t,s)) \qquad (۹-۱)$$

که $C_{j,t,s} \in \{0,1\}$.

تابع توزیع احتمال فوق از یک مجموعه داده آموزشی یادگیری می‌شود. داده‌های آموزشی شامل نمونه‌های آموزشی برچسب خورده هستند. هر نمونه آموزشی به صورت یک جفت از بردار ویژگی (نماینده هر کلمه) و برچسب مربوط به آن است. بعد از آموزش، توزیع احتمال یادگیری شده برای انجام طبقه بندی به کار می‌رود.

$$\hat{c} : (j,t,s) \rightarrow \hat{c}(j,t,s) \in \{0,1\} \qquad (۱۰-۱)$$

یک برچسب بر اساس فرمول زیر به یک بردار نسبت داده می‌شود:

$$\hat{c}(j,t,s) = \underset{c\in\{0,1\}}{\operatorname{argmax}}\, p(c;j,t,s) \qquad (۱۱-۱)$$

در عبارت بالا برچسبی به کلمه اختصاص می‌یابد که احتمال آن از ۰/۵ بیشتر باشد ولی این مقدار حد آستانه، گاهی نامناسب است چرا که هزینه دادن برچسب نادرست به کلمه درست بسیار کمتر از هزینه دادن برچسب درست به کلمه نادرست است. برای همین یک حد آستانه τ تعریف می‌شود:

$$\hat{c}(j,t,s;\tau) = \begin{cases} 1 & if\ p(1;j,t,s) > \tau \\ 0 & otherwise \end{cases} \qquad (۱۲-۱)$$

از عبارت بالا برمی‌آید که اگر احتمال درستی یک کلمه از مقدار حد آستانه بیشتر باشد برچسب درست می‌خورد و از آنجا که در برخی کاربردها تشخیص کلمات نادرست مهمتر است، در این موارد مقدار حد آستانه بیشتر از ۰/۵ در نظر گرفته می‌شود. مقدار حد آستانه می‌تواند با استفاده از معیارهای ارزیابی روی مجموعه آموزشی بهینه شود.

به طور خلاصه انجام تخمین اطمینان در سطح کلمه نیاز به دو کار دارد: تولید ویژگی‌های مناسب و طبقه بندی مناسب.

۱-۳ طرح مسئله

اکثر فناوری‌های زبان طبیعی ناکاملند. یکی از دلایل آن این است که به طور کلی در بر گرفتن پیچیدگی‌ها و ابهامات زبان طبیعی سخت است. دلیل دیگر به استفاده از روش‌های آماری برمی‌گردد. روش‌های آماری در دو دهه اخیر کاربرد گسترده‌ای در تشخیص گفتار و پردازش زبان طبیعی داشته‌اند و به موفقیت‌های قابل ملاحظه‌ای نیز رسیده‌اند. این موفقیت به دلیل این واقعیت است که این رویکرد مستقل از زبان است و تنها به پیکره‌های به اندازه کافی بزرگ برای تخمین چگالی‌های احتمال نیاز است. با این حال در روش‌های آماری یک مشکل وجود دارد: آنها تنها محتمل ترین نتایج را با توجه به داده‌های آموزشی و ورودی تولید می‌کنند. روشن است که این روش گاهی با توجه به انتظار انسان بهینه نیست [12]. با این وجود می‌توان با استفاده از راه

حل‌هایی خطای موجود در فناوری‌های ناکامل را کنترل کرد. برای مثال در سامانه تشخیص گفتار، درجه اطمینان پایین در آنالیز سخن یک کاربر می‌تواند موجب شود سامانه عملیات را تکرار کند. این استراتژی قابلیت بهبود دادن کارایی سامانه را دارد، به شرطی که تخمین اطمینان دقیق باشد [10]. بنابراین سنجش کیفیت نتایج، به طور خودکار مهم است. تخمین اطمینان این کار را انجام می‌دهد.

با وجود اینکه از زمان ظهور ترجمه ماشینی یعنی تقریبا ۶۰ سال پیش، بهبودهای وسیعی در این حوزه به وجود آمده‌است، کارایی سامانه‌های ترجمه ماشینی مدرن نیز هنوز فاصله زیادی تا کامل شدن دارند. ترجمه یک کلمه می‌تواند اشتباه، جابجا، یا غیر موجود باشد. همچنین کلمات اضافه‌ای می‌توانند تولید شوند. با این حال سامانه‌های ترجمه ماشینی موجود برای کاربردهایی از جمله ترجمه سریع برای جستجو و یا کمک به مترجمان انسانی برای ترجمه مفید است [19]. برای کاربردهایی که فقط نیاز به مراد جمله مبدا دارند و نیاز به ترجمه دقیق ندارند، سامانه‌های ترجمه ماشینی جدید خصوصا سامانه‌های ترجمه ماشینی آماری جوابگو هستند. اما برای کاربردهایی از جمله ترجمه شرح اقدامات و مذاکرات رسمی که ترجمه دقیق مورد نیاز است، سامانه‌های ترجمه ماشینی فقط می‌توانند یک ترجمه ابتدایی ارائه دهند. سپس مترجم‌های انسانی این ترجمه‌های ابتدایی را با پس‌ویرایش تصحیح می‌کنند [20]. اما گاهی ممکن است پس‌ویرایش زمان بیشتری نسبت به ترجمه مستقیم توسط انسان بگیرد. به دلیل نبودن دانش در مورد کیفیت ترجمه اولیه، مترجمان انسانی ممکن است زمان خود را برای پس‌ویرایش ترجمه‌های بسیار بی ربط هدر دهند. در چنین حالتی، داشتن تخمینی از درستی یا اطمینان خروجی ترجمه ماشینی برای ویرایشگران مطلوب است تا تلاش خود را به جملاتی معطوف کنند که نیاز به تغییرات پرهزینه ندارند. همچنین این تخمین‌ها می‌توانند به ویرایشگران کمک کنند که بخش‌های درست ترجمه را پس‌ویرایش نکنند [20]. همچنین تشخیص خطا با تخصیص معیارهای اطمینان به هر کلمه، عبارت یا جمله ترجمه شده، برای بهبود کیفیت سامانه مفید است [12]. تخمین اطمینان سامانه را قادر می‌کند خطاهای محتمل را به اطلاع کاربر برساند یا تنها ترجمه‌های با اطمینان بالا را ارائه دهد [14].

در پاسخ به اینکه « چرا از امتیازاتی که خود سامانه پایه به خروجی‌ها می‌دهد برای درجه اطمینان استفاده نشود؟ » باید گفت استفاده از لایه تخمین اطمینان برروی سامانه پایه، نسبت به استفاده از امتیازاتی که خود سامانه پایه به خروجی اختصاص می‌دهد مزایایی دارد. مزیت کلیدی آن این است که به صورت یک ماژول جدا از سامانه است. این مسئله مزایای زیر را به دنبال دارد: