

## ارائه یک مدل تطابق جملات در ترجمه ماشینی مبتنی بر مثال با استفاده از الگوریتم ژنتیک

رضا اکبری<sup>۱</sup>، محمد هادی صدرالدینی<sup>۲</sup>، سید مصطفی فخر احمد<sup>۳</sup>

<sup>۱</sup> دانشجوی کارشناسی ارشد، دانشکده مهندسی برق و کامپیوتر، دانشگاه شیراز، شیراز  
reza.akbarie@gmail.com

<sup>۲</sup> عضو هیئت علمی، دانشکده مهندسی برق و کامپیوتر، دانشگاه شیراز، شیراز  
sadredin@shirazu.ac.ir

<sup>۳</sup> عضو هیئت علمی، دانشکده مهندسی برق و کامپیوتر، دانشگاه شیراز، شیراز  
mfakhrmahmad@cse.shirazu.ac.ir

### چکیده

در این مقاله، ما یک مدل را در جهت اندازه‌گیری میزان شباهت دو جمله در ترجمه ماشینی مبتنی بر مثال ارائه کرده‌ایم. در مدل ارائه شده از الگوریتم ژنتیک و یک تابع برازندگی جدید که مبتنی بر بار معنایی منطبق شده بین دو جمله می‌باشد استفاده گردیده است. ما فعل‌ها را به عنوان قلب یک جمله در نظر گرفته‌ایم چون بخش اساسی یک جمله بشمار می‌آیند و مقادیر زیادی از بار معنایی جمله را حمل می‌کنند. بنابراین ما در تابع برازندگی ارائه شده توجه بیشتر خود را بر روی افعال جمله قرار داده‌ایم. این نکته قابل توجه است که مدل ارائه شده دارای وابستگی زیادی به بخش گفتاری و مترادف‌های کلمات استخراج شده از وردنت و همچنین ترتیب کلمات دارد. نتایج بدست آمده از آزمایشات انجام شده بر روی دو پیکره (مجموعه مثال) نشان می‌دهد که الگوریتم ارائه شده، کیفیت تطابق جملات را در ترجمه ماشینی مبتنی بر مثال بهبود بخشیده است و ما به میانگین دقت ۷۹.۵ درصد در آزمایشات انجام شده بر روی پیکره‌ها رسیده‌ایم.

### کلمات کلیدی

ماشین ترجمه، ترجمه ماشینی مبتنی بر مثال، تشابه جملات، الگوریتم ژنتیک، بار معنایی، فاز تطابق.

### ۱- مقدمه

از معانی کلی هر دو جمله می‌توان "علاقه داشتن افراد به ورزش" را استنباط کرد که بیان‌کننده یک میزان تشابه قوی بین دو جمله است. اما اگر فاعل و مفعول در نظر گرفته شود، یک تفاوت قابل ملاحظه‌ای بین دو جمله وجود دارد که تفاوت افراد را به تفاوت ورزش‌ها ترجیح می‌دهد. بنابراین اگر دو جمله دارای فعل یکسان یا مترادفی باشند، به احتمال خیلی بالایی دارای مفهوم یکسان خواهند بود. اما در دو جمله بالا دیده می‌شود که سایر کلمات نیز باعث تفاوت‌های معنایی در جملات شده‌اند. ما مقدار ۱.۰ را به عنوان حداکثر و ۰.۰ را به عنوان حداقل مقدار میزان شباهت بین دو جمله در نظر می‌گیریم. در جملات، فعل‌ها نقش بسیار کلیدی نسبت به سایر کلمات ایفا می‌کنند. و با حذف افعال، جمله نامفهوم خواهد شد. بنابراین باید معنایی کلی جمله را نیز در نظر داشت. دو جمله مشابه هستند اگر بار معنایی یکی از آنها به بار معنایی دیگری نزدیک شود [۴]. ما نیز مانند [۵] نتوانستیم هیچ تعریف دقیق و روشنی از "بار معنایی" پیدا کنیم. ما میزان تشابه را با تعریف بار معنایی منطبق شده میان دو جمله بدست

ترجمه ماشینی مبتنی بر پیکره [۱] یکی از برجسته‌ترین زمینه‌های ترجمه ماشینی می‌باشند. در ترجمه ماشینی مبتنی بر مثال یک جمله جدید از زبان مبدأ با استفاده از مثال‌های ترجمه شده قبلی به جمله‌ای به زبان مقصد ترجمه می‌گردد [۲]. در این نوع از ماشین‌های ترجمه فاز تطابق، فرآیندی است که مشابه‌ترین جملات را برای یک جمله ورودی، از یک مجموعه مثال‌های از قبل آماده شده استخراج می‌کند. این فرآیندها مسلماً به یک مقیاس اندازه‌گیری تشابه نیازمندند [۳]. تاکنون مطالعات زیادی بر روی فرآیند تطابق توسط محققان انجام گردیده است. یک راه‌حل بلفوه که می‌توان برای تعریف معیار در نظر گرفت، وزن‌دهی به کلمات موجود در یک جمله می‌باشد. بر این اساس هر کلمه در جمله دارای وزن خاصی بوده که وزن آن با سایر کلمات می‌تواند متفاوت باشد. برای مثال دو جمله  $S1$  و  $S2$  را در نظر بگیرید:  
 $S1$  : Young People like playing football.  
 $S2$  : Old people like walking.