

اطلاعیه دفاع

نام دانشجو: سیدسروش مجد		نام استاد راهنما: دکتر مهرنوش شمس فرد	
مقطع: کارشناسی ارشد		رشته: مهندسی کامپیوتر	
نوع دفاع:		گرایش: هوش مصنوعی، رباتیک و رایانش شناختی	
<ul style="list-style-type: none"> <input type="checkbox"/> دفاع پروپوزال <input checked="" type="checkbox"/> دفاع پایان نامه <input type="checkbox"/> دفاع رساله دکترا 		تاریخ: ۱۴۰۳/۶/۲۶	
		ساعت: ۱۱-۱۳	
		مکان: ساختمان دانشکده مهندسی برق و کامپیوتر- اتاق ۱۱۷	
عنوان: بهبود عملکرد شباهت یابی معنایی متون بالینی به کمک یادگیری عمیق			
داوران خارجی: دکتر یدالله یعقوب زاده		داوران داخلی: دکتر حامد ملک	
<p>چکیده:</p> <p>شباهت‌یابی معنایی متون یکی از فناوری‌های پردازش زبان طبیعی است که به ارزیابی میزان هم‌ارزی معنایی دو قطعه متن می‌پردازد و نقشی حیاتی در بهبود بازیابی اطلاعات از منابع پزشکی، ادغام داده‌های ناهمگن و سازماندهی پرونده‌های الکترونیکی سلامت ایفا می‌کند. در این پژوهش برای بهبود شباهت‌یابی معنایی متون بالینی به حل چالش‌هایی نظیر کمبود و توزیع نامتوازن داده‌ها، مقاومت در برابر بازنویسی جملات، افزایش تعمیم‌پذیری می‌پردازیم. به این منظور دو رویکرد نوآورانه در یادگیری متریک عمیق برای بهبود بازنمایی مدل‌ها پیشنهاد کردیم. رویکرد پیشنهادی اول با روش ماسک‌دراپ‌اوت تلاش می‌کند تا مدل را با داده‌های بیشتری آموزش داده و عملکرد آن را بهبود بخشد. رویکرد پیشنهادی دوم از ترکیب دو تابع زبان تنظیم همبستگی پیرسون و رتبه‌بندی منفی چندتایی استفاده می‌کند. هدف این رویکرد نزدیک‌تر کردن جملات بازنویسی شده به جملات اصلی و دور کردن جملات نامرتبط از یکدیگر است که باعث بهبود درک معنایی مدل از جملات هم‌معنی با ساختارهای نحوی و واژگانی متفاوت می‌شود. همچنین برای شناسایی اصطلاحات پزشکی و جلوگیری از بیش‌برازش، مترادف‌های این اصطلاحات از هستان‌شناسی پزشکی MeSH استخراج و هنگام آموزش مدل در جملات اصلی جایگذاری شدند. برای رفع مشکل کمبود داده و ارزیابی بهتر رویکردها، با استفاده از مدل زبانی بزرگ لاما ۳.۱ و MeSH، مجموعه‌ای از جملات بازنویسی شده با درجات سختی متنوع از مجموعه داده ClinicalSTS تولید کردیم. بازنویسی‌ها شامل تغییراتی در ساختار نحوی، ترتیب کلمات و استفاده از مترادف‌ها هستند. روش پیشنهادی دوم و ادغام هستان‌شناسی MeSH، برای حالتی که یک جمله از جفت جمله ورودی و یا هر دو بازنویسی شوند، به ترتیب با کسب میانگین ضریب همبستگی پیرسون ۹۴.۴۱ و ۹۰.۶۳ به بالاترین مقاومت دست یافته و تعمیم‌پذیری قابل قبولی دارد. همچنین رویکرد استفاده از مترادف‌های موجود در MeSH منجر به تعمیم‌پذیری بالاتر اکثر رویکردها شده که از بین رویکردهای کدگذار دوقلو، تنظیم همبستگی پیرسون در مجموعه آزمون اصلی و بازنویسی شده به ترتیب با ضرایب همبستگی پیرسون ۸۹.۰۶ و ۸۶.۷۲ بهترین عملکرد را داشته است.</p>			