



دانشگاه اصفهان

دانشکده مهندسی کامپیوتر

گروه مهندسی نرم افزار

باسمه تعالی

جلسه دفاع از پایان نامه کارشناسی ارشد

مهندسی کامپیوتر گرایش مهندسی نرم افزار

تولید پاسخ در سیستم گفتگوی گفتاری هم دلانه

با استفاده از ویژگی های صوتی و متنی

سیستم های گفتگو در قالب های متنی و صوتی، نقش پر رنگی در زندگی روزمره ما دارند و در بسیاری از وظایف از جمله دستیار شخصی هوشمند، خدمات مشتری، مشاوره و غیره، مورد استفاده قرار می گیرند. به عبارت دیگر این سیستم ها نوعی رابط کاربری هستند و با انسان ها تعامل دارند. بنابراین مواردی چون تقویت درک از کاربر و پاسخگویی متناسب با آن می تواند به بهبود رابط کاربری منجر شود و در نتیجه رضایتمندی کاربران را افزایش دهد. توانایی درک احساسات دیگران و ارائه پاسخ متناسب با آن ها، در انسان ها به عنوان هم دلی شناخته می شود. به همین دلیل، افزودن هم دلی به سیستم های گفتگو به یکی از موضوعات مهم پژوهش های اخیر جهت بهبود این سیستم ها تبدیل شده است. در بسیاری از پژوهش ها، تنها از وجه متن گفتگو جهت افزودن هم دلی به سیستم های گفتگو استفاده شده است. این در حالی است که در سیستم گفتگوی گفتاری، گفتگو در قالب صوت انجام می شود و صوت حاوی اطلاعات زیادی مانند لحن، بلندی صدا، شدت، مکث، لرزش صدا و زیر و بمی است. این اطلاعات می تواند نشانگر سطح استرس، احساسات، عواطف، جنسیت و سن کاربر باشد و تنها بخشی از آن چیزی است که می توان از صوت استنتاج کرد. بر اساس پژوهش های روان شناسی، صوت می تواند در برانگیختن هم دلی نقش موثری ایفا کند. علاوه بر این، پژوهش های بسیاری نشان داده اند که ترکیب اطلاعات صوتی و متنی توانسته است عملکرد مدل های تشخیص عاطفه را بهبود بخشد. با وجود این موارد، تعداد بسیار کمی از پژوهش ها به استفاده از صوت برای ایجاد پاسخ های هم دلانه پرداخته اند. اغلب این پژوهش ها تنها اطلاعات محدودی از صوت را به صورت متنی استخراج کرده و سپس همراه با تاریخچه گفتگو به مدل های زبانی بزرگ ارائه داده اند که این روش منجر به نادیده گرفتن بسیاری از اطلاعات مهم و موثر موجود دیگر در صوت می شود. در این راستا، این پژوهش روشی جهت تولید پاسخ های هم دلانه با بهره گیری از ترکیب وجوه صوت و متن ارائه می دهد. در گام اول این پژوهش، به دلیل عدم وجود مجموعه داده ی گفتگوهای هم دلانه دو وجهی (متن و صوت) برای آموزش یک سیستم تولید پاسخ انتهابه انتها، مجموعه داده ای متناسب با نیازهای پژوهش با استفاده از خط لوله طراحی شده در این پژوهش، و اعمال آن بر روی چهار مجموعه داده ی گفتگوی چند وجهی تهیه شده است که این مجموعه داده با نام BiMEmpDialogues شناخته می شود. در ادامه یک مدل تولید پاسخ هم دلانه دو وجهی طراحی شده است که از دریچه ی متحرک جهت ادغام وجوه صوت و متن بهره می برد. این مدل مبتنی بر دانش خارجی و نمونه بوده و از سه طبقه بند تشخیص وجود سازوکارهای ارتباطی هم دلی در پاسخ، در هنگام آموزش استفاده شده است تا مدل را به سمت تولید پاسخ ایده آل هم دلانه هدایت کند. طبق ارزیابی های انجام شده، نسخه ی متنی مدل پیشنهادی دارای عملکرد خوبی نسبت به پژوهش های اخیر بوده و پاسخ های تولید شده توسط مدل، دارای امتیاز حضور هم دلی بالاتری می باشد. هم چنین مدل پیشنهادی از منظر معیارهایی چون BLEU و ROUGE و نیمه از ابعاد کیفیت معیار FED عملکرد بهتری نسبت به نسخه متنی خود نشان داده است.

استاد راهنما:

دکتر افسانه فاطمی

استاد داور داخل:

دکتر محمدعلی نعمت بخش

استاد داور خارج:

دکتر مریم حسینی

پژوهشگر:

زلفا شفرئی

تاریخ برگزاری:

۱۴۰۳/۱۱/۱

ساعت:

۱۴:۰۰

مکان برگزاری:

اتاق ویدئو کنفرانس