

چکیده

محبوبیت رسانه‌های اجتماعی و پذیرش وسیع آنها توسط کاربران در سرتاسر جهان باعث ایجاد یک منبع غنی از اطلاعات بدون ساختار جهت درک رفتار کاربران، علایق و نظرات آنها شده است. امروزه شناسایی رویدادهای مهم جهان از روی محتوی‌های تولید شده توسط کاربران، به عنوان یک حوزه فعال شناخته شده است. این رویدادها زندگی گروه بزرگی از افراد را تحت تاثیر قرار می‌دهد در حالی که یک فرم منحصر به فردتری از رویدادها بنام رویداد شخصی از اهمیت بالایی برخوردارند و برخلاف رویدادهای جهانی بر زندگی افراد کمی اثر می‌گذارند.

شناسایی رویدادهای شخصی مانند ازدواج، مسافرت، استخدام در مقایسه با سایر رویدادها مشکل‌تر می‌باشد، زیرا (۱) این رویدادها خاص یک کاربر می‌باشند و بازخورد وسیعی ندارند. (۲) این رویدادها به صورت مستقیم مطرح نمی‌شوند و نیاز به استنتاج از محتوی‌های تولید شده توسط کاربران می‌باشد. (۳) بسیاری از کاربران رویدادهای شخصی خود را بر روی رسانه‌های اجتماعی گزارش نمی‌کنند و همین امر سبب شده است که مسأله شناسایی رویدادهای شخصی یک مسأله با عدم تعادل بالا باشد.

هدف ما از این کار تحقیقاتی ارائه یک روش برای شناسایی رویدادهای شخصی با استفاده از داده‌های متنی رسانه‌های اجتماعی می‌باشد و با توجه به اینکه بخش زیادی از این داده‌های متنی در مورد رویدادهای آینده می‌باشد، توانمندی این روش در شناسایی رویدادهای شخصی که در آینده برای کاربر رخ می‌دهد از اهداف این تحقیق می‌باشد. اما از آنجایی که در مسأله شناسایی رویداد آینده کاربر به اصل تویییتی که در آن به رویداد شخصی اشاره شده است دسترسی وجود ندارد، باید مجموعه‌ای از تویییت‌ها که کاربر در گذشته منتشر کرده است را بکار برد که آن را تاریخچه کاربر می‌نامیم. حال اگر بتوان از تاریخچه‌های کاربران موجود، که برای همه‌ی آنها در آینده رویداد شخصی رخ خواهد داد، نشانه‌هایی مشترک یافت، می‌توان با کمک این نشانه‌ها مسأله شناسایی رویداد شخصی در آینده را حل کرد.

راهکار ارائه شده در این رساله برای یافتن این نشان‌های مشترک شامل دو بخش اصلی است. در ابتدا، با تکیه بر تکنیک تعبیه کلمه به منظور ارائه یک مدل معنایی از رویداد شخصی، رویدادهای شخصی کاربران از تویییت‌های گذشته آنها استخراج و در نتیجه تاریخچه کاربران از نظر رویداد شخصی ساخته می‌شود. سپس، دو رویکرد متفاوت بنام تولید محتوا و کشف روابط بین رویدادهای شخصی برای شناسایی رویداد شخصی آینده کاربران معرفی شده‌اند. رویکرد تولید محتوا که مبتنی بر شبکه عصبی بازگشتی است بازه کوتاه مدت از تویییت‌های گذشته کاربر را مورد پردازش قرار می‌دهد در حالی که رویکرد دوم با تکیه بر تکنیک‌های یادگیری ماشین، نشانه‌های مشترک را از بازه طولانی مدت از تویییت‌های گذشته کاربر کشف می‌کند.

نتایج حاصل از آزمایش‌ها نشان می‌دهد مسأله شناسایی رویداد شخصی آینده کاربر، یک مسأله غیربديهی است و راهکار پیشنهادی توانسته است به یک کارایی قابل قبولی بر روی مجموعه داده استاندارد طلایی برسد.