

تشخیص حالت فرد از روی متن

تا بحال نرم افزارهای زیادی برای تشخیص غلط های املایی یا گرامری یک متن منتشر شده اند و تحقیقات زیادی بر روی این مقوله انجام شده است، اما کمتر کسی به دنبال استخراج حالت نویسنده در هنگام نوشتن آن بوده است. حالت و احساسات درونی یک انسان بخشی از ارتباطات روزمره او است. حالات درونی یک فرد را می توان از برخی از رفتارهای خروجی او مانند اظهار نظرها، حالت چهره و رفتارهای وی تشخیص داد. امروزه علوم ثابت کرده اند که حالات درونی یک فرد را می توان از روی بسیاری از ویژگی های فیزیکی و جسمانی او نیز مشخص کرد. این حالات می توانند تاثیر مستقیمی بر روی فشار خون، قند خون، ترکیبات فرمولی و آنزیمی درون خون و حتی خروجی مواد شیمیایی از غدد ترشحی عرق و چربی از سطح پوست بگذارند. یکی دیگر از رفتارهای خروجی که نشان دهنده حالات درونی فرد است گرامر و نحوه نوشتاری او است. حالات عمومی درونی یک فرد را می توان به موارد "خوشحالی"، "ناراحتی"، "عصبانیت"، "تعجب"، "تنفر"، "وحشت"، "ترس" و "خنثی" تقسیم بندی کرد. همچنین می توان به هر یک از این حالات چهار سطح داد که این سطوح می توانند "زیاد"، "متوسط"، "کم" و "طبیعی" باشند. یک عامل انسانی به راحتی می تواند حالات درونی یک فرد را در هنگام نوشتن یک متن با خواندن نوشته هایش تشخیص دهد. تاثیر حالات درونی بر روی متن را می توان در حوزه روانشناسی و زبان شناسی قرار داد، لذا می توان در حوزه پردازش زبان طبیعی و یادگیری ماشین به بررسی و تحقیق در این مورد پرداخت و یک سیستم خودکار ماشینی تشخیص حالات درونی از روی متن را طراحی کرد. زبان یک ابزار قدرتمند برای برقراری ارتباط بین انسان ها است و از این رو تکنیک های پردازش زبان طبیعی (NLP) مدتهاست که برای شناسایی خودکار و استخراج اطلاعات از روی متن بکار می روند. اعمالی که (NLP) بر روی متن های نوشتاری انجام می دهد عبارت از طبقه بندی، خلاصه سازی، سیستم های پرسش و پاسخ، بازیابی اطلاعات از روی متن و تشخیص مود و حالت درونی فرد نویسنده هستند. در چند سال اخیر تحقیقات بر روی شناسایی حالت درونی از روی متن توسط تکنیک های هوش مصنوعی بسیار مورد توجه قرار گرفته شده است. تشخیص حالات درونی نیازمند پردازش های طبیعی زیادی از قبیل "ادراک"، "استدلال"، "یادگیری" و "پردازش زبان طبیعی" است.

پروفسور رنه دسکارتس (Rene Descartes) در رساله اش که در سال 1969 منتشر شد (احساسات روان) اولین تئوری ها را مبتنی بر تاثیر حالات درونی بر روی رفتارهای خروجی ارائه کرد که طبق نظریه وی حالات درونی به دو بخش حالات درونی اولیه و حالات درونی ثانویه تقسیم می شوند. پس از وی نیز تحقیقات زیادی در مورد ایجاد یک دسته بندی منظم و گویا از حالات درونی انجام شده و تعداد این دسته ها بیشتر شدند. برخی از متدهای پیشنهاد شده را در زیر بررسی می کنیم:

روش های مبتنی بر کلمات کلیدی

این گونه متدها از کلمات مترادف و متضاد موجود بر اساس ریشه نظری کلمه برای تعیین احساسات استفاده می کنند. از جمله کارهای انجام شده تئوری معرفی شده Liu & Hu بوده که یک روش خود راه انداز که از یک مجموعه کوچک ریشه نظری کلمات موجود در (wordnet.princeton.edu) WordNet برای یافتن کلمات مترادف و متضاد استفاده کرده است. اما این گونه روش ها سه نوع محدودیت می توانند داشته باشند:

1- ابهام در تعریف کلمات کلیدی

برخی از کلمات به کار برده شده خودشان دارای معنایی هستند که یک احساس را می رسانند، از این جمله می توان به کلمات "عجبا"، "احسنت"، "هورا"، ... اشاره کرد که اگر در ابتدای جمله ای که قرار داشته باشند احساس نهفته درون جمله را مشخص می کنند. اما کلمات بسیاری هستند که در موقعیت های مختلف احساسات متفاوت را بیان می کنند. مانند کلماتی که در جملات کنایه دار بکار می روند.

2- گاهی اوقات یک جمله حاوی یک کلمه کلیدی دارای حس خاص نیست ولی همان احساس را می رساند، مانند "من در آزمون پذیرفته شدم" و "هورا، من در آزمایش پذیرفته شدم".

3- جملاتی که دارای اطلاعات زبانی کمی هستند. از این قبیل جملات می‌توان به این دو جمله اشاره کرد که فقط در آنها جای دو ضمیر عوض شده اما کاملاً احساس متضادی را می‌رسانند، "من از روی تمسخر به او خندیدم" و "او از روی تمسخر به من خندید".

روش فضای برداری (VSM)

این روش از یک ماتریس استفاده می‌کند که سطرهای آن مشخص‌کننده کلمات و ستون‌های آن معرف جملات، پاراگراف‌ها و اسناد هستند.

روش‌های کلاسیفیکیشن

رویکردهای زیادی تاکنون از این روش استفاده کرده‌اند و بیشتر آنها بر این باورند که کلاسیفای کردن احساس و رفتار همان مشکلات کلاسیفای کردن متن را دارا است. در این نوع کلاسیفای کردن فرض بر این است که احساسی که از یک متن مورد انتظار است عبارت است از مجموع احساسات تمام کلمات سازنده آن. حال برخی از کلمات دارای مفهوم و احساس روشنی بوده و برخی دیگر دارای ابهام هستند. کارهایی که بر روی روش‌های کلاسیفیکیشن انجام گرفته است شامل تشخیص بخش موضوع اصلی متن، تشخیص اینکه آیا در متن می‌توان موضوعی احساسی یا رفتاری را یافت و اگر وجود دارد در کدام بخش است، یافتن جهت و سمت و سوی آن و درجه‌ای که می‌توان به آن داد نیز بسیار مهم و قابل توجه است. موضوعاتی که قبلاً بیشتر بر روی آن‌ها کار شده است نیز عبارتند از تشخیص موضوع احساسی فیلم‌ها، داستان‌ها، مقالات و مطالب مهم روزنامه‌ها و در این اواخر بیشتر مطالب وبلاگ‌ها بوده است.

یکی از مهمترین بخش‌ها در شناسایی حالت فرد از متن این است که بتوان از درون یک متن جملات حاوی مفاهیمی حاوی رفتار و احساس را درک کرده و آن جمله را از سایر جملات تشخیص داد و تفکیک کرد. این تشخیص نوعی کلاسیفای کردن است که متن را باید به دو کلاس جملات شامل حالات رفتاری و جملات فاقد حالات رفتاری تقسیم می‌شوند. برخی از روش‌های معرفی شده از شبکه‌های بیز و برخی از ماشین‌های بردار پشتیبان برای این کلاسیفای کردن استفاده کرده‌اند. از جمله رویکردهایی که می‌توانند در این حوزه تحقیقاتی مورد استفاده قرار بگیرند روش‌های مبتنی بر منطق فازی است. از آنجا که تشخیص حالات و احساسات یک فرد از روی هر خروجی او منجمله متن یک رفتار فازی بوده، بررسی و تشخیص مود نویسنده متن مسئله‌ای فازی است. همچنین از درون یک متن می‌توان حالات مختلفی را استنباط کرده اما یکی از آنها حالت درونی غالب بوده که رویکردهای فازی می‌توانند کمک بزرگی در این زمینه کرده و به نتیجه‌های مطلوبی دست پیدا کنند.