



تشخیص وضعیت بیمار بر اساس توضیحات ارائه شده‌ی متنی بیمار

پروپوزال پروژه‌ی نهایی مباحثی در علوم کامپیوتر

دکتر اکبری

پاییز ۹۹

امیرعلی کابلی - ۹۶۱۳۰۲۰

مهراد شاه‌محمدی - ۹۵۱۲۰۳۴

شیده هاشمیان - ۹۶۱۳۴۲۹

مقدمه

امروزه با توجه به همه‌گیری استفاده از سایت‌های دارویی که کاربران در آن تجربه و دلیل استفاده از دارویی را با یکدیگر به اشتراک می‌گذارند، محتوای متنی تولید شده توسط آن‌ها می‌تواند حاوی اطلاعات مفیدی باشد. این اطلاعات می‌تواند شامل دارویی مصرفی در شرایط خاص، عوارض داروها، محبوبیت یک دارو برای یک بیماری، شرایط بیمارانی که مریضی خاصی را دارند و... باشد. در این پروژه قصد بر این است که با استفاده از مجموعه داده‌ای که ویژگی گفته شده را دارد، بر روی یافتن حالت و وضعیت (دلیل استفاده نکردن از کلمه‌ی بیماری این است که این حالت می‌تواند سرگیجه یا چیزی شبیه به این باشد) اشخاص با توجه به توضیحاتی که داده‌اند تمرکز داشته باشد. نکته‌ی قابل توجه این است که این توضیحات افراد در زمان مصرف دارو است.

به بیانی دقیق‌تر هدف در این پروژه آموزش دسته‌بندی است که با گرفتن توضیحات فرد در حال مصرف دارو، شرایطی (یا بیماری‌ای) که شخص در آن قرار دارد را خروجی دهد. هدف اولیه پروژه -با توجه به تعداد خیلی بالای دسته‌ها (توضیحات بیشتر در بررسی داده آمده است)- تنها بررسی تعدادی از پرتکرارترین وضعیت‌ها (بیماری‌ها) است.

اما با توجه به جذابیت مسئله، در صورت امکان، سعی در پیش‌بینی نوع داروی مصرفی نیز خواهیم داشت.

جذابیت مسئله

پیش‌بینی کردن وضعیت یک فرد و بر اساس متن دریافتی، کاربردهای زیادی می‌تواند داشته باشد و امکانات جالبی در اختیار ما قرار دهد. (مثال‌هایی که بیان می‌شوند، امکاناتیست که مدل‌های کلی و پیشرفته‌تر مدل ما می‌توانند در زمینه‌ی پزشکی داشته باشند.)

به عنوان مثال:

- در صورت عدم دسترسی به دکتر، می‌توان به کمک شرح دادن وضعیت خود، ایده‌ای از وضعیتی که در آن قرار داریم، داشته باشیم.
- پیش‌بینی داروهای ممکن برای رفع وضعیت‌های موضعی بر اساس نسخه‌های تجویز دکتران
- چک کردن واقعی یا جعلی بودن داروهای تجویز شده‌ی یک نسخه، برای بیماری یا وضعیت موجود

مجموعه‌ی داده

مجموعه داده‌ی "[UCI ML Drug Review Dataset](#)" برای این پروژه در نظر گرفته شده است. این داده اولین بار توسط [UCI Machine Learning repository](#) منتشر شده است و با کراال کردن توضیحات آمده شده توسط مصرف‌کنندگان دارو، در سایت‌های دارویی جمع‌آوری شده است.

این مجموعه شامل ۱۱۲۳۲۹ داده‌ی آموزش و ۴۸۲۸۰ داده‌ی آزمایش است.

ستونهای این مجموعه شامل:

- شناسه (uniqueID)
- نام دارو (drugName)
- نام وضعیت (condition): وضعیت، حالت و یا بیماری‌ای که بیمار برای آن دارو مصرف شده است.
- توضیحات بیمار: متنی شامل وضعیت کاربر که توسط خود او در یک شبکه مربوط به اینگونه مسائل نوشته شده است.
- امتیاز شخص توضیح دهنده (rating): یک عدد بین ۰ تا ۱۰ که امتیاز درج کننده‌ی نظر است.
- تاریخ درج توضیحات (date)
- میزان کارایی (usefulCount): عددی بین ۰ تا ۹۴۹ که میزان کاربردی بودن آن پیام توسط دیگر کاربران است.

ارزیابی

ارزیابی مدل پیاده‌سازی شده، به کمک معیارهای معمول مسائل دسته‌بندی مانند موارد زیر انجام می‌شود:

- Accuracy
- Precision
- Recall
- F1-Score(micro and macro)
- Confusion Matrix

مراجع

نکته: مراجع آورده شده، نوع تعریف مسئله آنها شبیه و مرتبط با این مسئله است، اما تعریف دقیق آنها یکی نبوده و لزوماً مرجعی برای پیاده‌سازی ما نیستند.

- Tran, T., & Kavuluru, R. (2017). Predicting mental conditions based on “history of present illness” in psychiatric notes with deep neural networks. *Journal of biomedical informatics*, 75, S138-S148.
- Kim, J., Lee, J., Park, E., & Han, J. (2020). A deep learning model for detecting mental illness from user content on social media. *Scientific Reports*, 10(1), 1-6.
- Sheikhalishahi, S., Miotto, R., Dudley, J. T., Lavelli, A., Rinaldi, F., & Osmani, V. (2019). Natural language processing of clinical notes on chronic diseases: systematic review. *JMIR medical informatics*, 7(2), e12239.