Drug Review Sentiment Analysis

Myra Rust
Fall 2021
https://myrarust/github.io/Myra-Rust/

Domain(s): Pharmaceuticals / Semantic Analysis

Drugs.com. (n.d.). Drug Index A to Z. Drugs.com. Retrieved from https://www.drugs.com/drug_information.html.

Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

Felix Gräßer, Surya Kallumadi, Hagen Malberg, and Sebastian Zaunseder. 2018. Aspect-Based Sentiment Analysis of Drug Reviews Applying Cross-Domain and Cross-Data Learning. In Proceedings of the 2018 International Conference on Digital Health (DH '18). ACM, New York, NY, USA, 121-125. DOI: https://dl.acm.org/doi/10.1145/3194658.3194677.

Grissette, Hanane & Nfaoui, El Habib & Bahir, Adil. (2017). Sentiment Analysis Tool for Pharmaceutical Industry & Healthcare. Transactions on Machine Learning and Artificial Intelligence. 5. 10.14738/tmlai.54.3339. Retrieved from https://www.researchgate.net/publication/320002093_Sentiment_Analysis_Tool_for_Pharmaceutical_Industry_Healthcare.

Huster, K. (2020, June 5). Drug Review Insights: Learning Data Science. Medium. Retrieved from https://medium.com/swlh/drug-review-insights-learning-data-science-adae6493b6cd.

Korkontzelos, I., Nikfarjam, A., Shardlow, M., Sarker, A., Ananiadou, S., & Gonzalez, G. H. Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts. Journal of Biomedical Informatics, Volume 62, August 2016, Pages 148-158. Retrieved from https://www.sciencedirect.com/science/article/pii/S1532046416300508.

Kunwar, A., Harode, R., & Malik, S. (2020, April 20). DRAW – Drug Review Analysis Work. Medium. Retrieved from https://medium.com/sfu-cspmp/draw-drug-review-analysis-work-96212ed98941.

Liu S, Lee I. Extracting features with medical sentiment lexicon and position encoding for drug reviews. Health Inf Sci Syst. 2019 May 30;7(1):11. doi: 10.1007/s13755-019-0072-6. PMID: 31168364; PMCID: PMC6542915. Retrieved from https://pubmed.ncbi.nlm.nih.gov/31168364/.

Pierre, S. (2019, November 24). Patient Sentiment for Pharmaceutical Drugs from Twitter. Towards Data Science. Retrieved from https://towardsdatascience.com/extracting-patient-sentiment-for-pharmaceutical-drugs-from-twitter-2315870a0e3c.

Repustate. (n.d.) NLP in Healthcare & Pharma. Repustate. Retrieved from https://www.repustate.com/health-and-pharma-sentiment-analysis-and-text-analytics/.

Data:

I found a dataset on UCI Machine Learning repository that contains text reviews of drugs by users.

```
# Look at the data
train.head()
```

	Unnamed: 0	drugName	condition	review	rating	date	usefulCount
0	206461	Valsartan	Left Ventricular Dysfunction	"It has no side effect, I take it in combinati...	9.0	May 20, 2012	27
1	95260	Guanfacine	ADHD	"My son is halfway through his fourth week of ...	8.0	April 27, 2010	192
2	92703	Lybrel	Birth Control	"I used to take another oral contraceptive, wh...	5.0	December 14, 2009	17
3	138000	Ortho Evra	Birth Control	"This is my first time using any form of birth...	8.0	November 3, 2015	10
4	35696	Buprenorphine / naloxone	Opiate Dependence	"Suboxone has completely turned my life around...	9.0	November 27, 2016	37

Feature descriptions:

- drugName – the name of the drug being reviewed (categorical)
- condition – the reason(s) the drug is taken (categorical)
- review – Text review written by the patient (string)
- rating – rating provided by the patient on a 0-10 scale (numeric)
- date – date of review (datetime)
- usefulCount – number of users who marked the review as useful (numeric)

The dataset is already split into training and test datasets. The training dataset contains 161,297 observations and the test dataset contains 53,766 observations. The datasets are tab-separated files and can be downloaded from the following link.

Dataset Link: https://archive.ics.uci.edu/ml/datasets/Drug+Review+Dataset+%28Drugs.com%29

Research Questions:

I plan to conduct sentiment analysis on the text portion of the drug reviews. At first, I will focus my sentiment analysis on determining whether a review is positive or negative. That information should align with the rating column which relays overall patient satisfaction.

Research Questions:

Can sentiment analysis be conducted on drug reviews to determine if the review is positive or negative overall?

Can sentiment analysis be conducted in a way to tell us sentiment by drug?

Can sentiment analysis be conducted in a way to tell us sentiment by drug and condition?

Method:

I plan to use machine learning and natural language processing to conduct opinion mining of unstructured text data and classify each review as positive, negative, or neutral. If that is successful, I would like to see if I can use aspect-based sentiment analysis to break it down further and determine sentiment analysis by drug and then by drug and condition. For example: maybe a particular drug has a positive sentiment for one condition, but a really negative sentiment for another condition.

Potential Issues:

The columns drugName and condition are categorical, but I have seen that these columns sometimes contain more than one category per observation. I will have to figure out a way to deal with this, whether it is to create multiple observations from one or not, I won't know until I look more into the data.

Another challenge might be sentiment when dealing with a technical domain like drugs, drug reactions, side effects, effectiveness, etc. I have to see if I can find a lexicon that aligns with the sentiment type words that might be contained in the text.

Once again, I'm reaching out of my comfort zone and trying something new, so that always creates the potential for getting off schedule.

Concluding Remarks

It is important for an industry to be able to take reviews provided by clients and determine the overall sentiment of the product. The pharmaceutical industry is no different. Using machine learning and natural language processing, this project will conduct sentiment analysis on client reviews of drugs to determine the overall sentiment of the user review, the sentiment of the reviews by drug, and the sentiment of the reviews by drug and condition.