DL4H Final Project

Categorization of Free-text Drug Orders using Character-level Recurrent Neural Networks

Armaan R. Butt and Harikrishna Bojja
{arbutt2, hbojja2}@illinois.edu

Group ID: 213, Paper ID: 283

Code link: https://github.com/c3armaanbutt/dl4h_final_project

Agenda

- Overview of the Paper
- Scope of Reproducibility
- Data Profiling & Analysis
- ML Pipeline
- Results
- Conclusion

Overview of the **Original Paper**: Motivation

The annotation of active ingredients of free-text drug order entries is currently a labor intensive task. Standardization of drug order entries can be useful in clinical decision support for a patient. The researchers wanted to investigate the use of deep learning techniques to annotated the drug order entries. [1]

Overview of the **Original Paper**: Results & Conclusions

The researchers were able to achieve an accuracy on **96% using the baseline Trigram SVM model** and their **Bidirectional GRU RNN achieved 98%** accuracy. The feed-forward GRU, feed-forward & bidirectional LSTM models achieved similar performance to the baseline SVM.

Conclusion: Deep learning neural networks, specifically the bidirectional GRU, out performed current techniques and can help annotate free-text drug entries in EHR systems.[1]

Overview of the **Original Paper**: Dataset & Models

- The researchers used 26111 order entries from the University of Zurich Hospital System and manually labeled 2028 entries with 568 distinct drug codes (ATC).
- They employed fuzzy matching techniques (Jaccard Similarity) to automatically annotate the remaining records.
- The team trained a baseline trigram SVM model as well as feed-forward and bidirectional LSTM and GRU RNNs.[1]

Scope of Reproducibility

- The hypothesis that we wanted to show was that the **GRU RNN can perform better than the baseline SVM** as was shown in the paper.
- We reproduced the baseline trigram SVM model as well as the bidirectional GRU RNN with three hidden layer configurations (32, 64, 128).
- **4 distinct model configurations** were used with a total of 25 models trained **(22 SVM models and 3 RNNs)**
- We trained the SVM model on **22 drug labels** on **50000 records** and the GRU RNN for **binary classification** on **50000 records**.

Data Analysis: Initial Profile

Datasets: **PRESCRIPTIONS.csv** and **NOTEEVENTS.csv** from the MIMIC-III database.

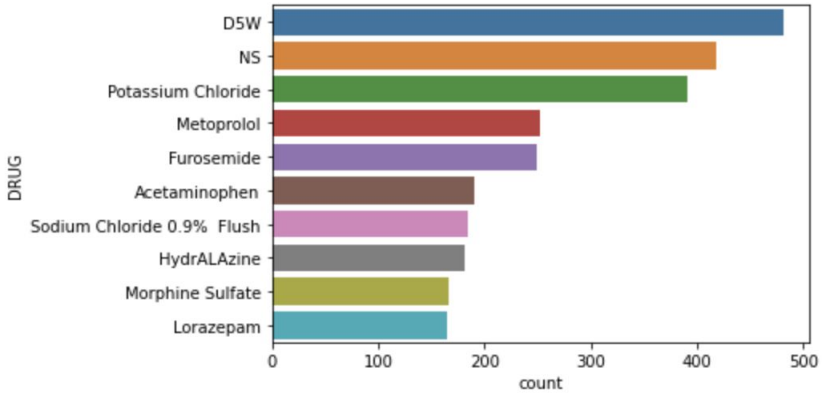
Number of Patients: 13033

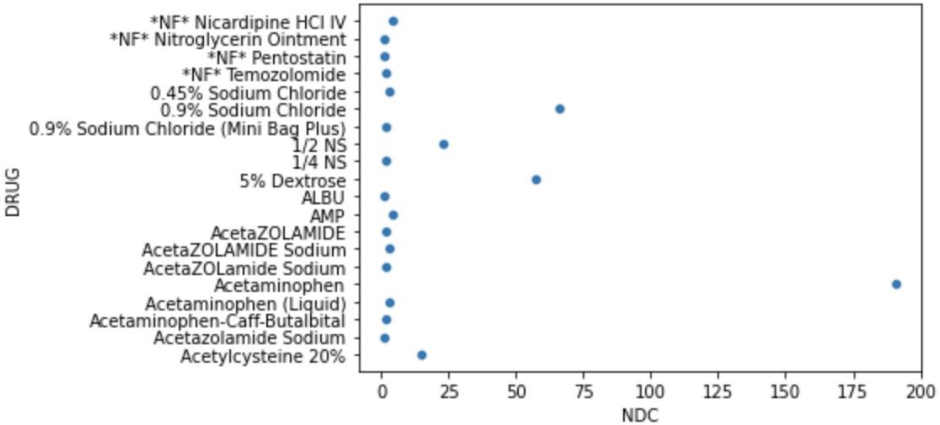
Total Number of Events: 50000

Unique Drugs: 4525

Unique NDC codes: 4204

Exploratory Data Analysis: Prescriptions





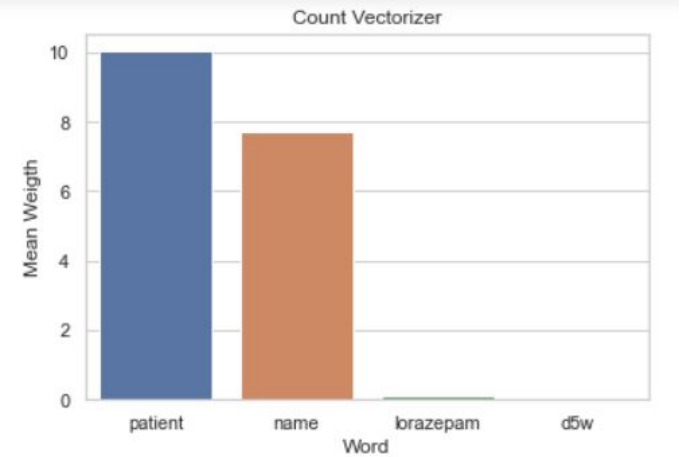
Drug Occurrences Conclusion:

- Prescriptions would have drugs that are prescribed hundreds of times
- Pick up the top n prescribed as there is better chances that these drugs would be a candidate for classification in corpus.

Drug to NDC Mappings Conclusion:

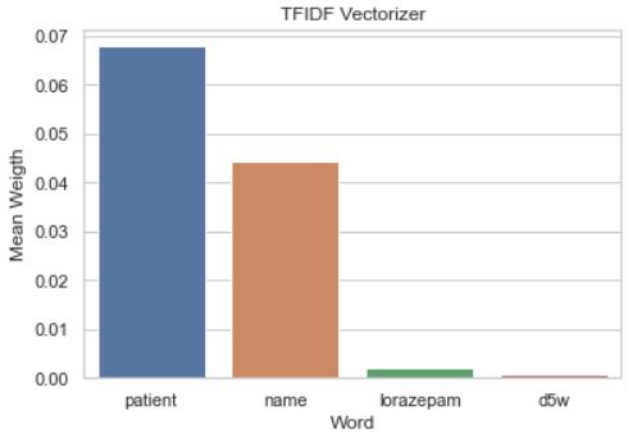
- A Drug could be mapped with more than one NDC code. One to many relation
- Pick the only one unique drug code to avoid ambiguity while training the model

Exploratory Data Analysis: Note Events



Count Vectorizer Feature Extraction:

- Count vectrorizer give weight to tokens/words based on number of occurrences
- Top 10 highest weighted words are not drugs
- Current classification problem may not be a best fit



TFIDF Vectorizer Feature Extraction:

- TFIDF vectrorizer assigns better weights to rarely occurred words
- We could evidently observe that drugs that are infrequent in corpus are assigned with better mean weight
- Current classification problem TFIDF Vectorizer seems to be better fit compared to count vectorizer

ML Pipeline Summary

EDA (Exploratory Data Analysis)

- Found multiple NDC codes per drug in prescriptions.
- Extracted 22 drugs and corresponding uniques to train.
- Extracted 50K events out of 2M events.
- Events have ~14K features with tri-grams and a lot of noise

Feature Extraction

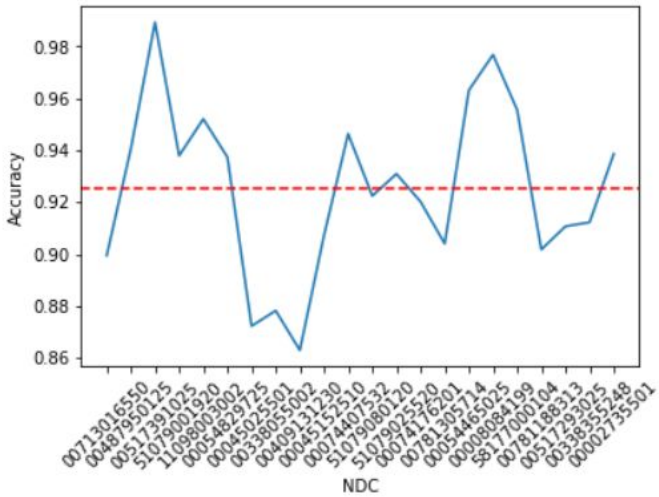
- Used **TFIDF vectorization to force higher weights** drug names in events.
- Applied **TrucatedSVD** to reduce the training data dimensionality and noise.

Model Training

- Training accuracy was low with 2.7G event data.
- Found **TFIDF Vectorization + TruncatedSVD** produced better results and train faster.

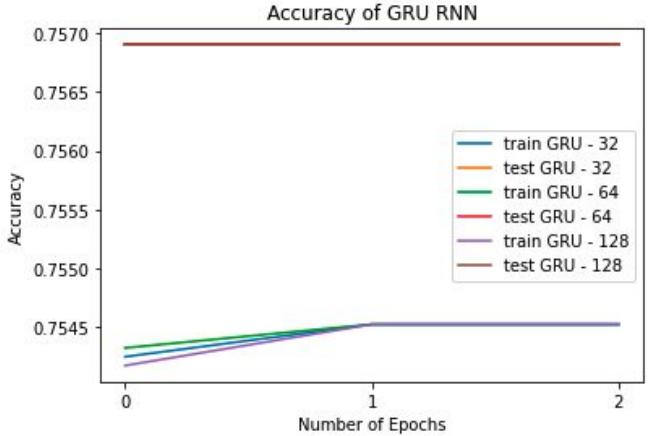
Results: SVM (Baseline)

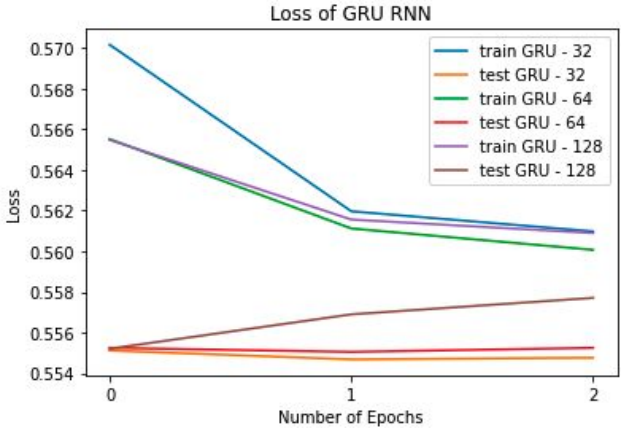
- Initial results of baseline model SVM trained on individual codes yielded an average accuracy of ~80%.
- Performance **jumped to 92%** after applying TFIDF and TruncatedSVD.



Results: Bidirectional GRU RNN

- Initial results with GRU RNN were 20%.
- Our data was more noisy than the original papers data set.
- Performance **jumped to 75%** after applying TFIDF and TruncatedSVD.
- Performance did not change significantly after multiple epochs.





Conclusion

- The original paper used the prescription data which has less noise than our clinical notes data.
- The paper reference that the GRU RNN was more susceptible to noise when compared to the SVM model and our project showed this [1].
- **To apply an RNN to clinical notes and not just drug orders, we need to use dimensionality reduction to yield better results.**
- Our team was able to reproduce the performance of the baseline SVM model but the performance of the GRU RNN is still behind claimed results.
- The event data that we contains lot of noise overfitting the dataset. We had to to improve the pipeline to include Trigram TFIDF and Truncated SVD.
- **SVM is more robust than GRU RNNs when the input data is has noise.**

References

[1] Raiskin Y, Eickhoff C, Beeler PE. Categorization of free-text drug orders using character-level recurrent neural networks. Int J Med Inform. 2019 Sep;129:20-28. doi: 10.1016/j.ijmedinf.2019.05.020. Epub 2019 May 23. PMID: 31445256.

Thank For Watching!

- We did not employee fuzzy matching techniques like the research team did and opted to annotate records based on an exact drug name.
- We used NDC instead of ATC as the output labels for each record.