VIRTUAL

SAS® GLOBAL FORUM 2021

#SASGF

Deep Learning to Classify Adverse Events from Patient Narratives

Qais Hatim, FDA CDER

Tom Sabo, SAS Institute Inc.

Qais is an expert in industrial engineering, advanced statistical modeling, machine learning, data analytics, and operational research. Qais is a data scientist at the U.S. Food and Drug Administration (FDA), where he performs work that has substantial merit and national importance.

Tom is a principal solutions architect at SAS who, since 2005, has been immersed in the field of text analytics and AI as it applies government challenges. This includes public health analytics, assessing research trends, and counter human trafficking solutions.

Key Points

Question: Can deep learning improve the efficiency of identifying safety signals from patient narratives?

Findings : In this study of 14,976 narrative observations in clinical study reports where 50% of such reports include indications of serotonin syndrome, the most successful deep learning method tested achieved a 98.88% correct classification rate of serotonin syndrome. Furthermore, machine learning methods which provide a layer of interpretability including associated symptoms achieved a 94.4% correct classification rate.

Meaning: Deep learning and machine learning can improve the speed, accuracy, and interpretability of medical coding for adverse events.

Introduction

- This is collaboration between SAS and FDA on the FDA Adverse Event Reporting System (FAERS)
- It contains detailed free-text narratives on adverse events occurring to a patient/subject
- Manual review and coding of these adverse events is hugely time consuming
- Automated coding of adverse events will improve postmarket and premarket safety reviews of FDA regulated drugs

Approach

- We’ve applied text analytics/ML in the past with success
- This initiative would leverage DL to classify one such event, *serotonin syndrome*, and could subsequently be leveraged for many such events
- We leveraged 4 different DL methods (tmCoOccur, tmCoOccur averaging, GloVe, Topic Weights) alongside an ML method which provides a layer of interpretability

Seratonin Syndrome

Target Variable under focus

Overview

Print

Serotonin syndrome occurs when you take medications that cause high levels of the chemical serotonin to accumulate in your body.

Serotonin is a chemical your body produces that's needed for your nerve cells and brain to function. But too much serotonin causes signs and symptoms that can range from mild (shivering and diarrhea) to severe (muscle rigidity, fever and seizures). Severe serotonin syndrome can cause death if not treated.

Serotonin syndrome can occur when you increase the dose of certain medications or add a new drug to your regimen. Some illegal drugs and dietary supplements also are associated with serotonin syndrome.

Milder forms of serotonin syndrome may go away within a day of stopping the medications that cause symptoms and, sometimes, after taking drugs that block serotonin.

Source: Mayo Clinic website. For informational purposes only

Symptoms

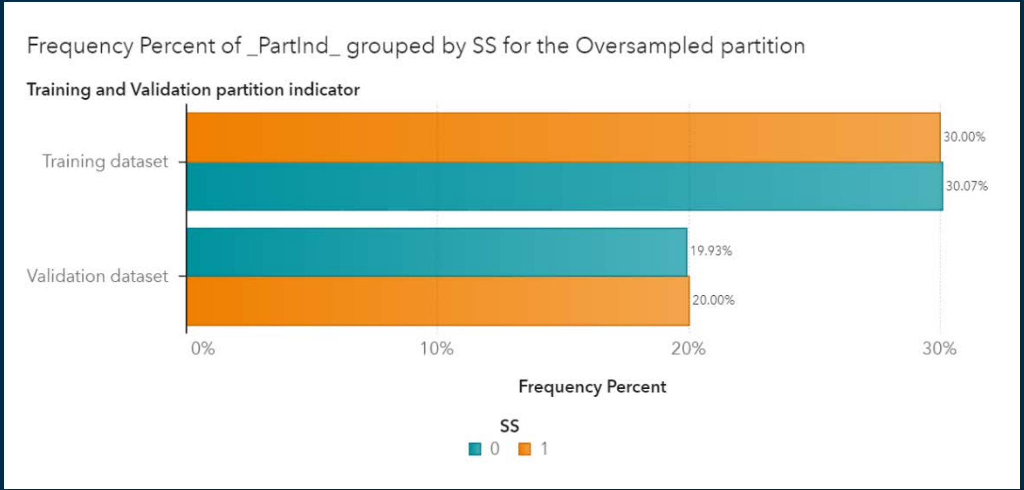
Serotonin syndrome symptoms usually occur within several hours of taking a new drug or increasing the dose of a drug you're already taking.

Signs and symptoms include:

- Agitation or restlessness
- Confusion
- Rapid heart rate and high blood pressure
- Dilated pupils
- Loss of muscle coordination or twitching muscles
- Muscle rigidity
- Heavy sweating
- Diarrhea
- Headache
- Shivering
- Goose bumps

Data Prep & Pre-Processing

- Extracted drug safety reports from the FDA system
- Flagged narratives with a Serotonin Syndrome flag (SS=1/0)
  - Drugs in one of two lists that had some likelihood of serotonin syndrome in the first place
- Oversampled for a 50% SS population, and a 60-40 training validation split



Takeaway 1: Use ALL the tools in your toolkit

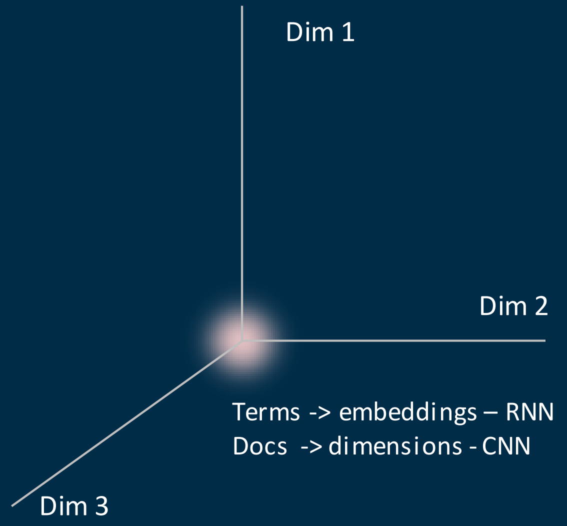
Word Embeddings from tmCoOccur

1. tmCoOccur embeddings

- Leveraged methods from Jim Cox's and Russell Albright's work
- Leveraged a sentence level window rather than a 3-5 word sliding window
- Applied CAS tmMine, tmCooccur, tmSvd, and an RNN

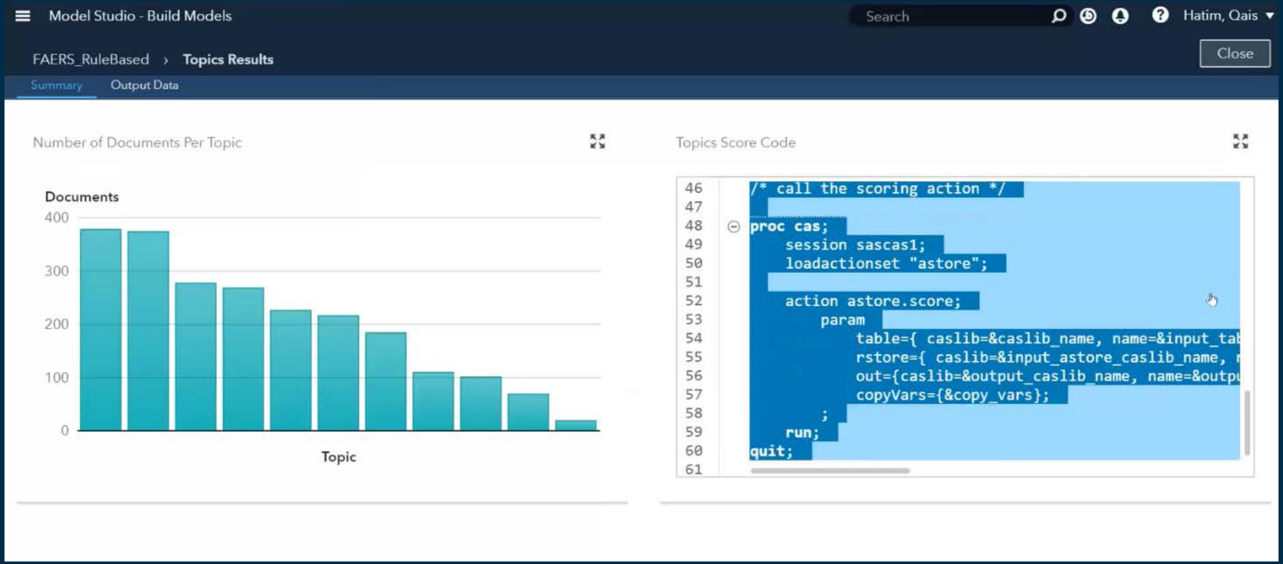
2. tmCoOccur averaged dimensions – direct inputs

- Additional step : Projected Co-occurrences directly on 200 dimensions and averaged the dimensions per document.
- Applied a CNN for modeling



Deep Learning with Topic Weights

- 3. Taking a leaf from Jim and Russell's paper, we examined if a document level approach outperforms a term embeddings and RNN approach.
- Used VTA topics, leveraged the topic weights for each training narrative as input to a CNN



#SASGF

SAS GLOBAL FORUM 2021

Takeaway 2: Customized embeddings prove better than off-the-shelf

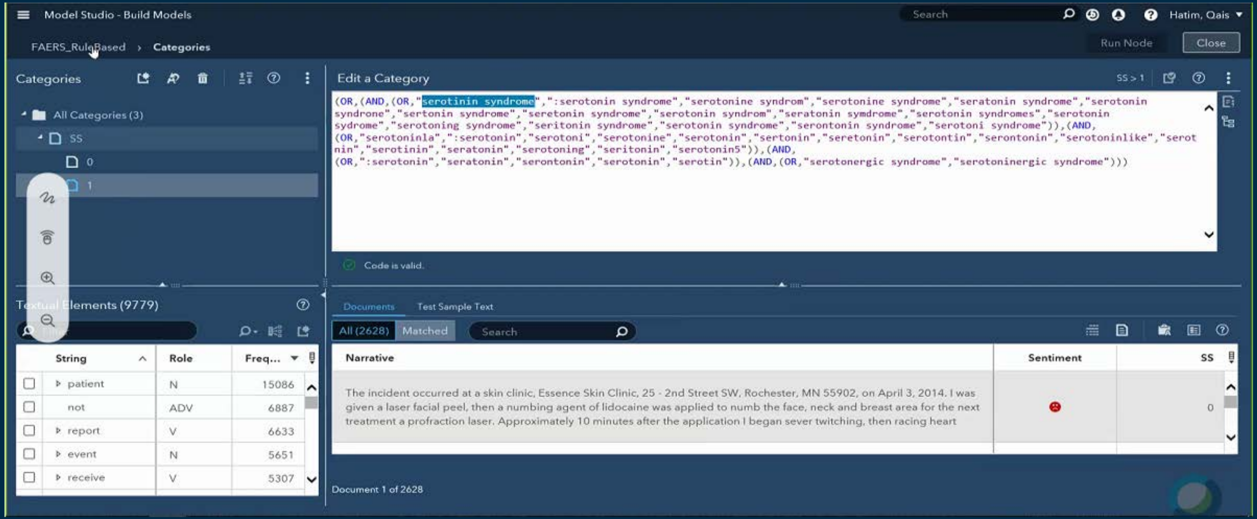
TmCooccur and GloVe

- 4. GloVe
  - Used standard GloVe 100-dim and 300-dim embeddings. Applied a RNN
- RNN
  - Used an RNN for both GloVe and TmCoOccur methods – a GRU model
  - Tuned through hyperparameter tuning
  - A long challenging process to get the right model!

	tmCooccur model	GloVe 300-dimention model
Model Type	Recurrent Neural Network	Recurrent Neural Network
Number of Layers	4	7
Number of Input Layers	1	1
Number of Output Layers	1	1
Number of Convolutional Layers	0	0
Number of Pooling Layers	0	0
Number of Fully Connected Layers	0	0
Number of Recurrent Layers	2	5
Number of Weight Parameters	101632	250496
Number of Bias Parameters	386	962
Total Number of Model Parameters	102018	251458
Approximate Memory Cost for Training (MB)	2469	5542

Takeaway 3: Follow a hybrid approach to text analytics

ML Boolean Rules Approach



- Took 5 iterations, leveraged an FDA SME in order to refine the model and identify new patterns
- Goal to generate a model that identified cases which indicate serotonin syndrome but had not been classified as such with keywords in the narrative.

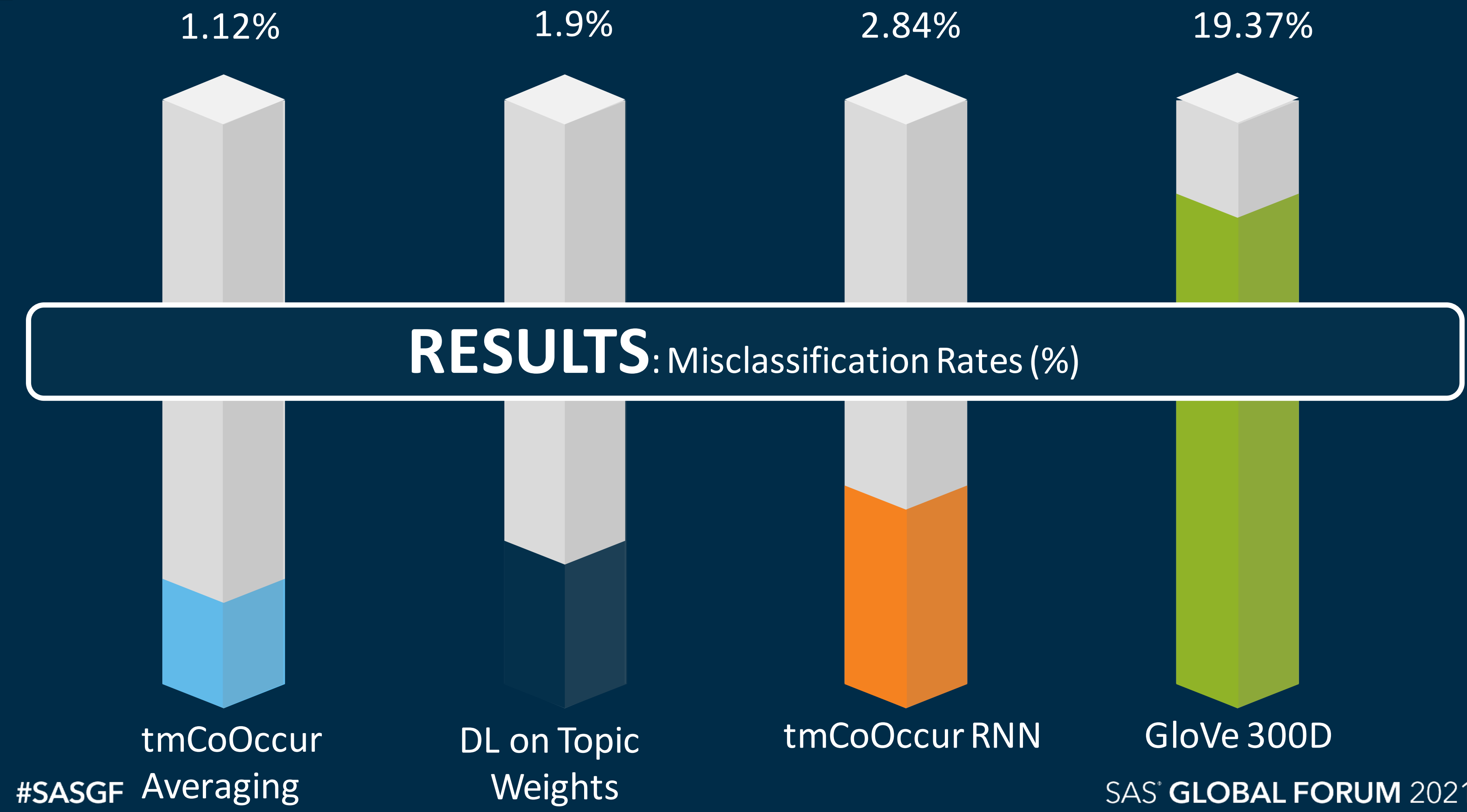
ML Boolean Rules Approach

Ruleset	Misclassification Rate
#1 (rules autogenerated after removing references in narrative to serotonin syndrome)	5.6%
#2 (rules from #1 + rules generated after an expert review involving additional stop words)	8.3%
#3 (rules from #1 + #2 rules generated after a second expert review involving additional stop words)	10.4%
#4 (rules from #1 - #3 + rules generated after a third expert review involving additional stop words)	12.8%
#5 (rules from #1 - #4 + rules generated after a third expert review involving additional stop words)	14.5%

- Rules provided a layer of interpretability and validation around the decision making process

Rule	_Target_	Average of _F1_	Average of _Precision_	Average of _Recall_
depression & tremor & tachycardia	1	0.930030088	0.892822026	0.970474282
syndrome	1	0.929925698	0.893195521	0.969806279
drug toxicity & venlafaxine	1	0.929803035	0.899875	0.961790247
symptom & hydrochloride & neuroleptic	1	0.929655707	0.899949975	0.961389446
temperature & mydriasis	1	0.929508303	0.900025025	0.960988644
syndrome & hydrochloride & hallucination	1	0.929360822	0.90010015	0.960587842
syndrome & ~attorney & ~site & ~initial information & ~arthritis & ~n	1	0.928899528	0.900526844	0.959118237
monoamine oxidase inhibitor	1	0.917918268	0.96193265	0.877755511
nms	1	0.917767065	0.9619215	0.87748831
suicidal & icu	1	0.917505593	0.962311189	0.876686707
hydrochloride & seizure & selective	1	0.917266942	0.962430291	0.876152305
icu & flush	1	0.917039731	0.962413743	0.875751503
toxicity & flush	1	0.916812426	0.96239718	0.875350701
syndrome & ~attorney & ~site & ~initial information & ~arthritis & ~n	1	0.916585025	0.962380603	0.8749499
syndrome & ~attorney & ~site & ~initial information & ~arthritis & ~p	1	0.91489511	0.963356974	0.871075484
lithium & ~consumer & major	1	0.910649756	0.969523235	0.858517034
linezolid & mental	1	0.910250957	0.969642048	0.857715431
anxiety & hyperthermia	1	0.909941852	0.969623697	0.857181029
anxiety & reaction & tremor & insomnia	1	0.909632572	0.969605323	0.856646627
anxiety & tremor & lorazepam	1	0.909168323	0.969577721	0.855845023
anxiety & jerk	1	0.908781146	0.96955468	0.855177021
escitalopram & disorient	1	0.908238636	0.969522365	0.854241817
toxicity & shake	1	0.907928389	0.969503869	0.853707415
tachycardia & ~attorney & diaphoresis & seizure	1	0.907617965	0.96948535	0.853173013

RESULTS: Misclassification Rates (%)



Additional Discussion Points

- This approach can be scalable and is production-ready for health conditions other than serotonin syndrome such as drug induced liver injury and cardiovascular cases
- Use DL in parallel with Boolean rule approach for an ensemble model; where models disagree, flag for manual review and possible misclassification
- We can apply BERT and BioBERT to leverage the best of pre-trained and customized embeddings

Thank you!

Contact Information

Qais.Hatim@fda.hhs.gov
tom.sabo@sas.com

VIRTUAL

SAS® GLOBAL FORUM 2021

#SASGF