OpenFDA Adverse Drug Prescriptions & Reactions Prediction Model

ARI KAMLANI

arikamlani.com

akamlani@gmail.com, ari.kamlani@nagra.com, @akamlani

http://github.com/akamlani/openfda

Problem

- Can we help Patients avoid adverse reactions **?**
- Minimized Reactions from Multiple Prescribed drugs
- Drug Avoidance based on High Probability of Reaction

Case Study

- Serious Reports
- Elderly Patients (Ages 65+), Patient Indication="Hypertension"
- Multiple Prescribed Drugs -> Multiple Reactions
- Determine Drug Correlation and Drug/Reaction Dependency

OpenFDA Queries

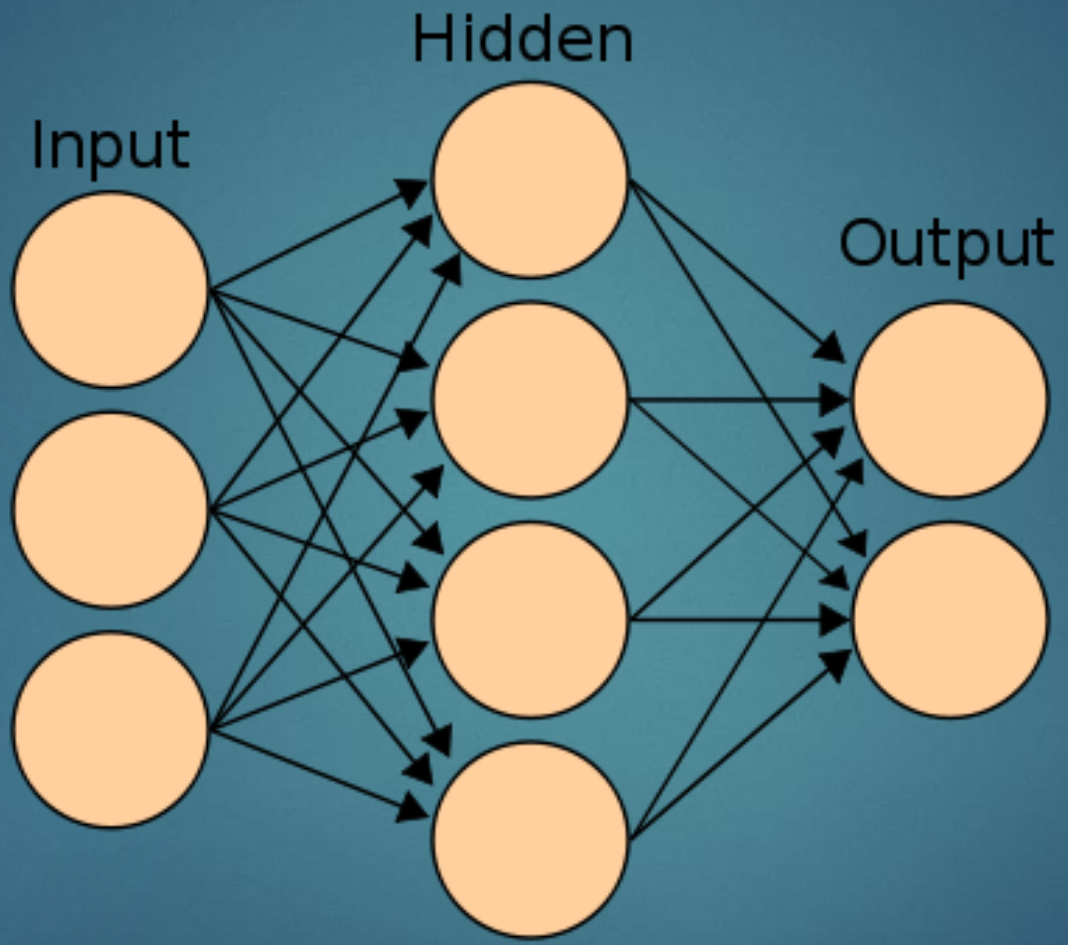
- OpenFDA Launched in Beta Summer 2014
- Drug Endpoint Query
  - https://api.fda.gov/drug/event.json?
  - Parameters: api_key, search, count, limit
  - Requires Pagination and Queries Rate Limited
- 4M + Reported Records
- Some records include Drug Name Normalization
- Via Voluntary Reports (FDA Adverse Event Reporting System: FAERS)
- JSON data not Flat (Multiple Embedded Dictionaries and Lists)
  - Not Able to Filter Multiple Levels (e.g. all Drugs, Reactions)

ML Techniques Used

- Multi-Dimensional Feature Space, Multiple Possible Targets
  - No True Target Label -> Semi-Supervised/Unsupervised Learning
  - Common ML Techniques learned don't fit
- Data Munging/Filtering – Voluntary Data
- Neural Networks Variants (Deep Learning)
  - Restricted Boltzmann Machine (RBM) – Binary Valued
- Cosine Similarity Metrics

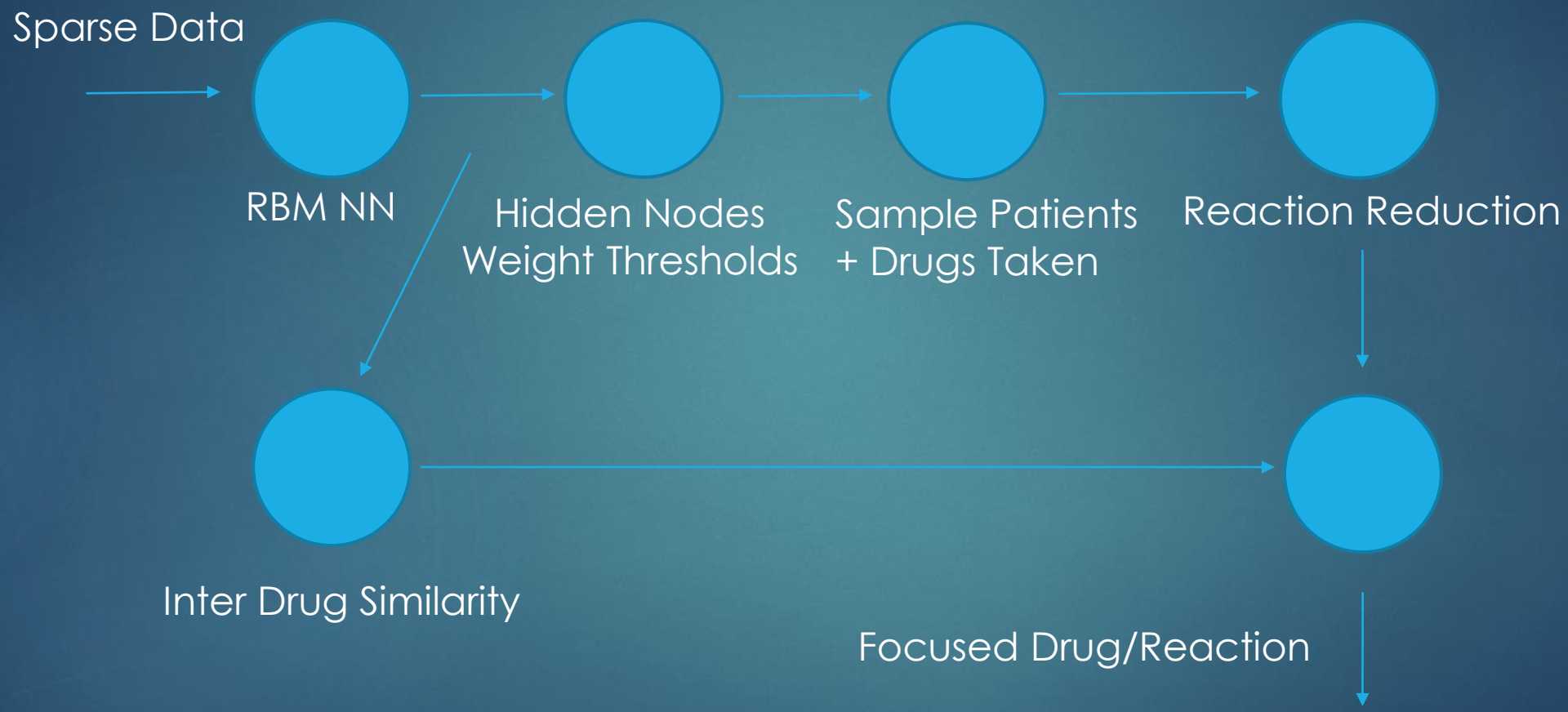
Neural Network Mappings

n_visible:
{d1, d2, d3, … dn}



n_hidden: n_reactions

Model


Sparse Data
RBM NN
Hidden Nodes
Weight Thresholds
Sample Patients
+ Drugs Taken
Reaction Reduction
Inter Drug Similarity
Focused Drug/Reaction

Stats

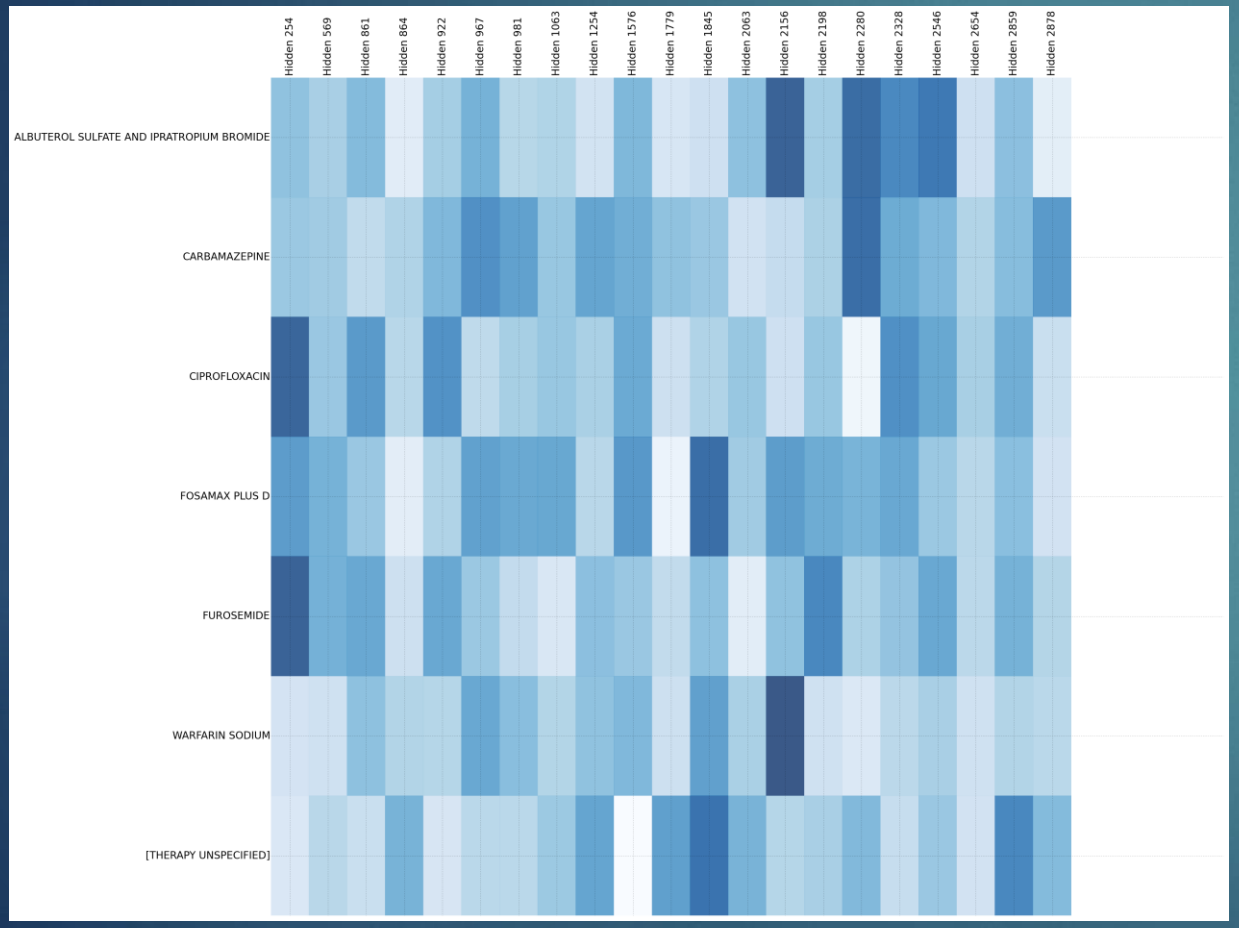
- 3000 Patient Training Records : Minimized due to Memory
- 4970 Visible Nodes (Drug Matrix)
- 2908 Hidden Nodes (Reactions)
- Default Number of Epochs (10)
- Training Time: < 10 Min
- Hidden Node Connection Threshold: >= 2

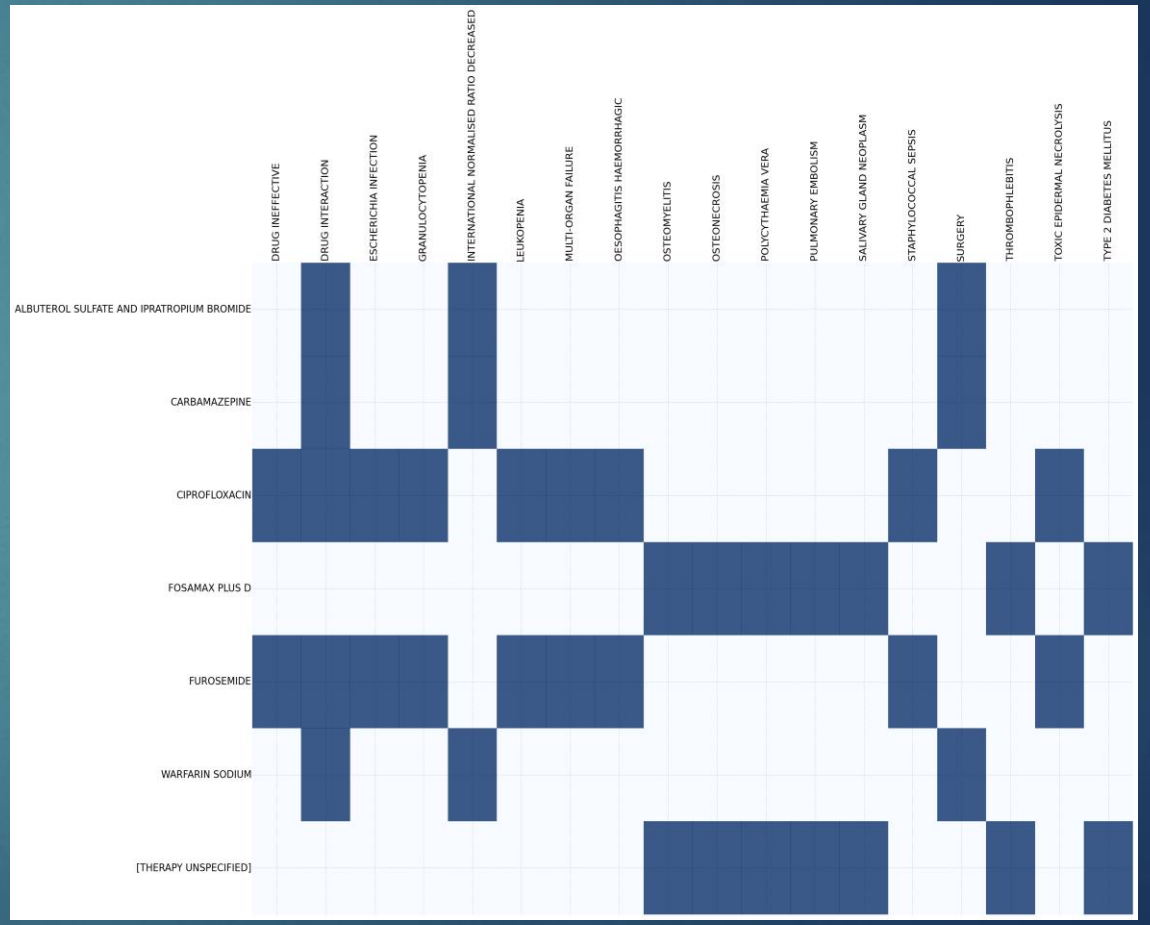
Top Similar Drugs Sample

Drug
ACCUZIDE (HYDROCHLOROTHIAZIDE, QUINAPRIL)
ERYTHROCIN
FOLBEE PLUS
FUSIDATE SODIUM
METFORMINE
NOLVADEX D
HERBESSOR R
SEISHOKU (SODIUM CHLORIDE) (SODIUM CHLORIDE)
BINORAC
PROPRANOLOL

Visualization:

Minimized Reactions from Drug Connection Weights





Lessons Learned

- Choose a simpler dataset and Project Idea!
- Non-Tabular data format creates more difficulty
- Dealing with Multidimensional data (Input + Output) is difficult
- Even small Neural Networks can take a long time to train
- Sparse Matrixes and Memory are important!
- Minimization problem to apply more traditional ML Aspects

Next Steps

- Use of Drug Harmonization
- Further Drug Separation and Insights
  - Chemical Structure
  - Physiologic Effect
  - Dosage Information
- Neural Networks
  - Theano Package
- Scaling
  - Apply to other Case Studies and Larger datasets
  - Multiple Clusters
  - Account for Memory

Q/A