UC San Diego School of Medicine

Department of BioMedical Informatics

Kevin Ngo Dr. Cinnamon Bloss, Dr. Jejo Koola

Using Deep Learning to Assist in Coding Qualitative Data

This research was supported by grant T15LM011271





- Upcoming 4th year at UCSD
- Majoring in Computer Science with Bioinformatics
- Desire to learn more about biomedical informatics
- Research Interest in:
 - Precision Medicine
 - Natural Language Processing
 - Machine Translation





Outline

- Deep Learning
- Word2vec
- Metrics
- Model Architecture
- Internal Validation Performance- DARPA Response
- Calibration
- External Validation Performance- FDA Comments



Deep Learning





What Is A Neural Network



What Is A Neural Network



ron

This r

8/22/2019







Word2Vec





Word2Vec



Japan – Tokyo + Russia = Moscow









Important Metric





Important Metric



Goal: Correctly Labeling Disease vs Not Disease

- Models:
 - Convolutional Neural Network (CNN)
 - Log. Reg. with TF-IDF (Justin)
 - Log. Reg. with Word Embedding
- Internally validated using bootstrapping with DARPA responses
- Recalibrated
- Externally validated with FDA comments
- Primary Metrics: AUC ROC, Sensitivity, Specificity

DARPA funds UC gene drive research against mosquito-borne diseases

ng female Anopheles gambiae mosquito, a known carrier of malaria. (CDC / James Gathany)





Methods



Model Architecture





Result

Model Comparsion



DARPA - ROC Curve



| | H (High) | M (Medium) | O (Youden's Point) | L (Low) |
|-------------|----------|------------|--------------------|---------|
| Sensitivity | 0.90 | 0.85 | 0.81 | 0.75 |
| Specificity | 0.63 | 0.75 | 0.80 | 0.86 |





Comparsion of Performance at Different Thresholds

Predicted Label



This research was supported by grant T15LM011271



Key Takeaways

- No Literary Works On Applying Deep Learning on Qualitative Data
- Newer ≠ Better
 - Word Embedding
- Different Situation = Different Thresholds



Lessons Learned

- Process of qualitative coding and analysis
- Neural Networks
- Word Embedding
- Calibration and Recalibration
- How to read machine learning papers

UC San Diego School of Medicine

Department of BioMedical Informatics

Acknowledgements

- Dr. Jejo Koola
- Dr. Cinnamon Bloss
- Dr. Cynthia Schairer
- Justin Castro
- The Bloss Lab

