Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for EHR Analysis

2018
Benjamin Shickel, Patrick J Tighe, Azra Bihorac, Parisa Rashidi
University of Florida

Presenter: Derrick Blakely
@ https://qdata.github.io/deep2Read/
Roadmap

1. Background
2. Motivation
3. Survey of Recent Advances in “Deep EHR”
4. Future Directions
Roadmap

1. Background
2. Motivation
3. Survey of Recent Advances in “Deep EHR”
4. Future Directions
Background - Electronic Health Records (EHRs)

- Huge increases in the numbers of EHRs in the US in the last 10 years
- Heterogeneous data
Background

- Huge increases in the numbers of EHRs in the US in the last 10 years
- Heterogeneous data
- Hospital admission and discharge data (datetime objects)
- Lab tests/results
- Radiological images
Background

- Huge increases in the numbers of EHRs in the US in the last 10 years
- Heterogeneous data
- Hospital admission and discharge data (datetime objects)
- Lab tests/results
- Radiological images
- Genomic data
- ICD codes
- Clinical notes (free text)
EHRs

- Primary use: bookkeeping, hospital administration
EHRs

● Primary use: bookkeeping, hospital administration
● Secondary uses:
  ○ Medical concept extraction
  ○ Patient trajectory modeling
  ○ Disease inference
  ○ Clinical decision support systems
  ○ Deidentiﬁcation
  ○ Phenotyping
EHR Analysis

- Traditional: logistic regression, random forests, SVM
EHR Analysis

- Traditional: logistic regression, random forests, SVM
- Recent: MLP, Autoencoder, RBM, Deep Belief Nets, CNN, RNN, GRU, and LSTM
EHR Analysis

- Traditional: logistic regression, random forests, SVM
- Recent: MLP, Autoencoder, RBM, Deep Belief Nets, CNN, RNN, GRU, and LSTM
- Most “Deep EHR” papers published in last 3 years
  - Several hundred total
Roadmap

1. Background
2. Motivation
3. Survey of Recent Advances in “Deep EHR”
4. Future Directions
Roadmap

1. Background
2. Motivation
3. Survey of Recent Advances in “Deep EHR”
4. Future Directions
Motivation

- Catalog advances
- High-level overview of what’s been going on in EHR analysis in the last few years
- Future directions
Roadmap

1. Background
2. Motivation
3. Survey of Recent Advances in “Deep EHR”
4. Future Directions
Roadmap

1. Background
2. Motivation
3. Survey of Recent Advances in “Deep EHR”
4. Future Directions
Deep EHR Overview

![Bar graph showing the number of publications for Deep EHR from 2012 to 2017. The number of publications increased significantly in 2015, 2016, and 2017.](image-url)
Deep EHR Overview

Deep EHR

Number of publications

Deep EHR: Application Areas
- Representation
- Representation learning
- Concept representation
- Phenotyping
- Information extraction
- Prediction
- Deidentification

Number of publications
Deep EHR Overview

Deep EHR: Technical Methods

- Unsupervised
- RNN
- LSTM
- GRU
- CNN
- Autoencoder
- RBM
- DBN
- Skip-gram

Number of publications

2012 2013 2014 2015 2016 2017
# Deep EHR Overview

<table>
<thead>
<tr>
<th>Task</th>
<th>Subtasks</th>
<th>Input Data</th>
<th>Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Extraction</td>
<td>(1) Single Concept Extraction</td>
<td>Clinical Notes</td>
<td>LSTM, Bi-LSTM, GRU, CNN RNN + Word Embedding</td>
</tr>
<tr>
<td></td>
<td>(2) Temporal Event Extraction</td>
<td></td>
<td>AE</td>
</tr>
<tr>
<td></td>
<td>(3) Relation Extraction</td>
<td></td>
<td>Custom Word Embedding</td>
</tr>
<tr>
<td></td>
<td>(4) Abbreviation Expansion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Representation Learning</td>
<td>(1) Concept Representation</td>
<td>Medical Codes</td>
<td>RBM, Skip-gram, AE, LSTM</td>
</tr>
<tr>
<td></td>
<td>(2) Patient Representation</td>
<td></td>
<td>RBM, Skip-gram, GRU, CNN, AE</td>
</tr>
<tr>
<td>Outcome Prediction</td>
<td>(1) Static Prediction</td>
<td>Mixed</td>
<td>AE, LSTM, RBM, DBN</td>
</tr>
<tr>
<td></td>
<td>(2) Temporal Prediction</td>
<td></td>
<td>LSTM</td>
</tr>
<tr>
<td>Phenotyping</td>
<td>(1) New Phenotype Discovery</td>
<td>Mixed</td>
<td>AE, LSTM, RBM, DBN</td>
</tr>
<tr>
<td></td>
<td>(2) Improving Existing Definitions</td>
<td></td>
<td>LSTM</td>
</tr>
<tr>
<td>De-identification</td>
<td>Clinical text de-identification</td>
<td>Clinical Notes</td>
<td>Bi-LSTM, RNN + Word Embedding</td>
</tr>
</tbody>
</table>
Roadmap

1. Background

2. Motivation

3. Survey of Recent Advances in “Deep EHR”

4. Future Directions
Roadmap

1. Background
2. Motivation
3. Survey of Recent Advances in “Deep EHR”
4. Future Directions
Future Directions: Representations

- Representations in light large amount of heterogeneity
Future Directions: Representations

- Representations in light of large amount of heterogeneity
- Heavy focus on clinical codes
Future Directions: Representations

- Representations in light large amount of heterogeneity
- Heavy focus on clinical codes
- Many things are not incorporated into representations/embeddings
Future Directions: Representations

- Representations in light large amount of heterogeneity
- Heavy focus on clinical codes
- Many things are not incorporated into representations/embeddings
- Clinical text is under-utilized
Future Directions: Representations

- Representations in light large amount of heterogeneity
- Heavy focus on clinical codes
- Many things are not incorporated into representations/embeddings
- Clinical text is under-utilized
- “Holy grail”: unified representation
Future Directions: Benchmarks

- Lack of universal benchmarks
- Difficult reproducibility
- Everyone claims “state-of-the-art performance”
- Proprietary data sets
- Hyperparameters can make or break an algorithm
Future Directions: Interpretability

- Models need to be transparent and trustworthy
- Explored so far: maximum activation, clustering illustrations, word clouds, heat maps, “Mimic learning”