Deep Learning Solutions for Classifying Patients on Opioid Use

S118: Detecting and Examining Patterns in Data: Applications for Combating Opioid Abuse and Adverse Events

Zhengping Che¹, Jennifer St. Sauver², Hongfang Liu², Yan Liu¹

¹University of Southern California
²Mayo Clinic
Disclosure

All authors have no relevant relationships with commercial interests to disclose.
Opioid Epidemic

• Opioids: Painkiller with high usage
  • 7% adults in US took opioids in last 30 days

• Opioid epidemic: “A public health emergency”
  • 91 Americans die everyday due to opioid overdose
  • Lots of consequences for patients and the public

• Important questions
  • What are the factors that contribute to opioid use?
    • Medical-related factors? Patient information?
  • What kind of patients will suffer from opioid misuse?
    • Accurate prediction and proper intervention?
    • Prediction and intervention at early stage?

In 2015...
- 2.1 million People misuse prescription opioids
- 15,281 People die from overdose on opioids
- 828,000 People use heroin
- 2.0 million People had prescription opioid use disorder
- 135,000 People use heroin for the first time
- 12,989 Deaths attributed to overdose on synthetic opioids

Opioid Usage Study

- Existing work on opioid usage study
  - [Hooten et al., 2015; Clarke et al., 2014; Thielke et al., 2017; etc]

- Data: Static features only (e.g., demographics, diagnoses/drug use history)
- Large-scale temporal EHR data from multiple chart tables

- Task: Identifying risk factors and indicators
- Classifying dependent users, making early prediction, identifying factors

- Method: Multivariate logistic regression & p-value test
- Powerful and scalable deep learning models & model investigations

Deep learning for opioid usage study is promising!
Deep Learning for Health Care

- Deep Computational Phenotyping
  - [Che et al., KDD 2015]

- Interpretable Deep Models
  - [Che et al., AMIA 2016]

- Deep Models for EHRs with Missing Values
  - [Che et al., arXiv 2016]

- Anomaly Detection
  - [Nilanon et al., CinC 2016]

- Variational Adversarial Deep Domain Adaptation
  - [Purushotham et al., ICLR 2016]
Data & Task

• Data source: Rochester Epidemiology Project (REP)
  • > 140,000 people
  • Covered 98.7% of the population in Olmsted County, MN
  • > 50 years of history

• Two classification tasks
  • Predict whether an opioid user will become a long-term user or short-term user
  • Predict if a long-term opioid user will be opioid-dependent
  • Identify factors and indicators in these two tasks

• Preliminary study [Hooten et al., 2015]
  • Same data source, ~300 patients, 1-year record, characteristics exploration
Data Processing – Cohort Selection

• Obtain outpatient drug prescriptions
  • Mayo Clinic & Olmsted Medical Center
  • Study period: Jan 1th, 2003 to March 31, 2016
  • Define analgesic prescriptions by RxNorm 2016 version
    • NDF-RT class code C8834 (Opioid Analgesics), or
    • Ingredient code 10689 (Tramadol) or 352362 (Acetaminophen/Tramadol)
  • Handle duplicated prescription data
    • Only keep the last prescription if the same drug is prescribed within 30 mins

• Determine cohorts
  • With authorization on the use of medical records
  • Ensure clean label
    • Exclude patients with prescriptions 6 months prior to first prescriptions within study period
  • N = 102,166 patients
Data Processing – Group Identification

• Label all patients into 3 groups
  • Short-term user (ST) and Long-term user (LT)
    • Similar definition used by the CONSORT study
    • Long-term user:
      • Episodes of opioid prescriptions >90 days, and
      • With >120 days supply or with >10 prescriptions
  • Opioid-dependent user (OD)
    • Subset of LT patients
    • Identified by the diagnosis of "opioid dependence" from their problem lists by Mayo NLP group
    • Not all OD patients are explicitly diagnosed.
  • Data characteristics matches preliminary studies

<table>
<thead>
<tr>
<th></th>
<th>Short-Term (ST)</th>
<th>Long-Term (LT)</th>
<th>Opioid-Dependent (OD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
<td>80,596</td>
<td>21,570</td>
<td>749</td>
</tr>
<tr>
<td>%</td>
<td>78.89%</td>
<td>21.11%</td>
<td>3.47%</td>
</tr>
<tr>
<td>% of LT</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Method – Feature Extraction

- **Features from 3 chart tables**
  - Diagnosis records (\textbf{DX}): ICD-9, ICD-10, HICDA code
  - Procedure and service records (\textbf{PR}): ICD-9, ICD-10, CPT/HCPCS code
  - Prescription records (\textbf{RX}): RxNorm ingredient code
  - Features to use: \textit{time step}, \textit{code}, and \textit{duration} (for \textbf{RX} only).

- **Feature mapping** is required
  - Raw code records are sparse
  - Coding system are different for different time/sites

<table>
<thead>
<tr>
<th>Table</th>
<th>DX</th>
<th>PR</th>
<th>RX</th>
</tr>
</thead>
<tbody>
<tr>
<td># of records</td>
<td>56,229,157</td>
<td>46,386,740</td>
<td>8,102,477</td>
</tr>
<tr>
<td># of raw codes</td>
<td>43,438</td>
<td>18,984</td>
<td>2,460</td>
</tr>
<tr>
<td># of mapped codes</td>
<td>284</td>
<td>245</td>
<td>307</td>
</tr>
</tbody>
</table>

\[ \Rightarrow \text{CCS code} \]
\[ \Rightarrow \text{CCS code} \]
\[ \Rightarrow \text{NDF-RT class code} \]
Method – Handling Temporal Data

- Make extracted features suitable for models
- **1-of-K (one-hot) encoding**
  - Convert each record into a fixed-length vector of size D
  - A prescription of \( m \) days is converted to \([0, \ldots, m, \ldots, 0]\)
- **Temporal sum-pooling** *Get fixed-length input*
  - Sum up all records of one patient into a vector of size D
  - Data can be used in any model
- **Temporal segmentation** *Keep temporal info*
  - Take sum-pooling for every year and stack them into a matrix
  - Data matrix is of size \( T \times D \), where \( T \) may be different
  - Data can be used in RNN model
Method – Deep Learning Predictive Models

- Two different deep learning models
  DNN: Multilayer feedforward network
  \[ x^{(l)} = f(W^{(l)}x^{(l-1)} + b^{(l)}) \]
  \[ y = x^{(L)} \]

  RNN: Recurrent neural network with Long Short-Term Memory (LSTM)
  \[ c_t, h_t = f(x_{t-1}, h_{t-1}, c_{t-1}) \]
  \[ x_{nn, t} = c_t \]
  \[ y = \sigma(Wx_{nn, T} + b) \]
Method – Other Baselines

• Three commonly used models are compared
  • Logistic Regression (LR)
  • Support Vector Machine (SVM)
  • Random Forest (RF)

• Implementation details
  • L1 penalty is applied in LR and SVM with hyperparameter tuning
  • Default setting of RF is used to avoid overfitting
  • Implemented in Python with scikit-learn package

• Can train and test efficiently on a desktop (no need for GPU!)

<table>
<thead>
<tr>
<th>Model</th>
<th>DNN</th>
<th>RNN</th>
<th>LR</th>
<th>SVM</th>
<th>RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Size (KB)</td>
<td>1,878</td>
<td>9,320</td>
<td>21</td>
<td>23</td>
<td>2,282</td>
</tr>
</tbody>
</table>
Discussion – Experiment Settings

- Three simulated settings
  - **Setting A**: Take all medical records before the labeling time
    - Ideal case with most input features and best prediction performance
  - **Setting B**: same as Setting A, excluding analgesics prescriptions
    - Try to find hidden indicators / factors
  - **Setting C**: Take all medical records before the first opioid prescription
    - Demonstrate early prediction capacity
Discussion – ST-LT Results

- **Setting A**
  - [Graphs showing performance metrics for different models in Setting A]

- **Setting B**
  - [Graphs showing performance metrics for different models in Setting B]

- **Setting C**
  - [Graphs showing performance metrics for different models in Setting C]
Discussion – LT-OD Results

- **Setting A**
- **Setting B**
- **Setting C**
Discussion – Feature Investigation Method

• Equation of prediction results in DNN

\[ y = \sigma(W^{(L)}f^{(L-1)}(...W^{(2)}f^{(1)}(W^{(1)}x + b^{(1)}) + b^{(2)}...) + b^{(L)}) \]

• Estimation on input feature importance in DNN

\[ I = W^{(L-1)}Bn^{(L-1)}(...W^{(2)}Bn^{(2)}(W^{(1)}Bn^{(1)})) \in \mathbb{R}^{1 \times D} \]

• \(d\)-th number in \(I\) is the importance score for \(d\)-th input feature
• Absolute value indicates the strength
• Sign indicates whether it’s positively or negatively relevant
• This is only a rough estimation, without considering
  • Feature dependencies in data
  • Bias term and non-linearity in models
### Discussion – Feature Investigation Results

#### Table / Code | Features for ST-LT Task | $I$
---|---|---
RX C8834 | Opioid Analgesics | 0.2287
RX C8890 | Amphetamine-like Stimulants | 0.0843
RX C8838 | Non-opioid Analgesics | 0.0802
PR CCS 227 | Other Diagnostic Procedures | 0.0272
DX CCS 258 | Other Screening | 0.0218
RX C4859 | Salicylates, Antirheumatic | 0.0204
DX CCS 203 | Osteoarthritis | 0.0185
DX CCS 205 | Spondylosis | 0.0179
DX CCS 98 | Essential Hypertension | 0.0126

#### Table / Code | Features for LT-OD Task | $I$
---|---|---
RX C8834 | Opioid Analgesics | 0.7784
DX CCS 661 | Substance-related Disorders | 0.6186
PR CCS 182 | Mammography | 0.3481
DX CCS 663 | Substance Abuse/Mental Health History | 0.3248
DX CCS 258 | Other Screening | 0.2948
PR CCS 228 | Prophylactic Vaccinations/Inoculations | 0.2796
DX CCS 651 | Anxiety Disorders | 0.2785
RX C8864 | Anticonvulsants | 0.2626
RX C8860 | Benzodiazepine Derivatives | 0.2382

---

**Note:** The table above lists the features that were found to be significant in the ST-LT and LT-OD tasks. The $I$ values indicate the importance of each feature in the task.
Summary

• Summary
  • A deep learning solution for opioid use study with promising results
  • Showcase of applying deep learning models in practical clinical studies with large-scale data

• Future work
  • Handle incomplete data and noisy features / labels
  • Incorporate numerical and unstructured data
  • Further investigate important indicators / factors
  • Capture temporal evolutions of opioid users and help prescription making

• Contact
  • Mayo NLP Group, Medical Informatics
  • USC Melady Lab
Thank you!

Email me at: zche@usc.edu
**Reference**

- **Opioid usage study**

- **Deep learning for health care**
Terms

- **RxNorm**: US-specific terminology in medicine that contains all medications available on US market
- **CONSORT**: Consortium to Study Opioid Risks and Trends, supported by the National Institute of Drug Abuse
- **ICD-9/10**: International Classification of Diseases, 9th/10th version
- **HICDA**: Hospital International Classification of Diseases Adapted (Mayo adaptation of ICD-8)
- **CCS**: Clinical Classifications Software, a uniform and standardized coding system based on ICD-9 into a smaller number of clinically meaningful categories
- **NDF-RT**: National Drug File - Reference Terminology
- **Bn**: Batch normalization, a technique used in deep learning by normalizing the activations of neural network layer at each batch.