medExtractR: medication extraction from electronic health records

HANNAH WEEKS
VANDERBILT UNIVERSITY
PK/PD studies

- Pharmacokinetics (PK): how the body affects the drug
  - Estimate parameters to explain drug movement within the body

- Pharmacodynamics (PD): how the drug affects the body
  - Dose-response relationships, adverse events

Pharmacokinetic studies

- Feedback through blood concentration measurements
PK/PD studies

Critical information for PK/PD studies:
- Blood concentration measurements
- Dosing information
  - What dosage did the patient take?
  - How often was that dosage taken?
  - At what time was the most recent dose taken?

Often found as unstructured data in EHR (clinical notes)

“...Patient takes tacrolimus 1mg 2x/day...”
Information extraction

- How do we extract data?
- Natural language processing (NLP)
  - Using computers to understand human language
- Information extraction
  - NLP task that converts unstructured input to structured output

“...Patient takes tacrolimus 1mg 2x/day...”

<table>
<thead>
<tr>
<th>Drug name</th>
<th>tacrolimus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dose</td>
<td>1 mg</td>
</tr>
<tr>
<td>Frequency</td>
<td>2x/day</td>
</tr>
</tbody>
</table>

medExtractR

- Targeted approach to medication extraction – intended to be used on a drug within a dataset
- Customizable through function arguments or modification of source code
- Written in R
  - Widely used for data analysis
  - Available on CRAN
SYSTEM DESCRIPTION

Medication Entities

- **Drug name**
- **Strength**: Amount of an individual unit (pill)
- **Dose amount**: number of units taken
- **Dose**: dose given intake (equivalent to strength x dose amount)
- **Frequency**: how often dose is taken
- **Intake time**: relative time of day when dose is taken
- **Dose change**: keyword indicating if dose is an increase, decrease, etc.
- **Last dose**: time at which the last dose was taken
Medication Entities

- Frequency
  - Any expression in the dictionary that is also in the search window will be extracted
  - Can be regular expressions
    - E.g., ‘q\sday’ will match ‘q day’ or ‘q/day’
  - Default dictionaries: `data(freq_vals)`, `data(intaketime_vals)`, `data(dosechange_vals)`

Dictionary-based entities

- Frequency
Rule-based entities

- **Strength**
  - ‘Number unit’
  - Function argument (unit = ‘mg’)
- **Dose amount**
  - ‘# (pill | tablet | capsule)”
  - ‘take | takes | taking #’
  - ‘(#)’
- **Last dose**
- **Time expression**
  - # am/pm
  - Military time (e.g., 2100)
  - Modifier (e.g., 10 last night)
  - Window includes ‘last | took | taken’

medExtractR functionality

**Input**
- clinical note, drug names, tuning parameters

**Internal**
- find drug names
- create search window
- identify/extract drug entities

**Output**
- data frame with drug entities
medExtractR example

medExtractR(note, drug_names = c("tacrolimus", "prograf", "tac"), unit="mg", window_length=60, max_dist=2)

"Patient is on tacrolimus 1mg (2) bid – took at 6:30 pm, cellcept 1000mg bid, prednisone 5mg daily."

<table>
<thead>
<tr>
<th>Entity</th>
<th>Expression</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>DrugName</td>
<td>tacrolimus</td>
<td>15:25</td>
</tr>
<tr>
<td>Strength</td>
<td>1mg</td>
<td>26:29</td>
</tr>
<tr>
<td>DoseAmt</td>
<td>2</td>
<td>31:32</td>
</tr>
<tr>
<td>Frequency</td>
<td>bid</td>
<td>34:37</td>
</tr>
<tr>
<td>LastDose</td>
<td>6:30 pm</td>
<td>48:55</td>
</tr>
</tbody>
</table>
Data

From the Synthetic Derivative – Vanderbilt University de-identified EHR

- Development drugs
  - Tacrolimus and lamotrigine
  - 60 training notes, 50 test notes

- Test drug
  - Allopurinol
  - 110 test notes

Data: train/test set selection

\[
\begin{array}{c|c|c|c}
\text{[patient_id]} & \text{taking drug} & \text{patient_1} & \ldots & \text{patient_n} \\
\end{array}
\]
Data

[patient_id] taking drug

[date] lab_value_11 ... lab_value_1m ... lab_value_n1 ... lab_value_nm

[patient_id] taking drug

[date] lab_value_11 ... lab_value_1m ... lab_value_n1 ... lab_value_nm

[note_id] from [date]

note_1

note_1

... note_k

note_1

note_1

note_2

note_1
Sample randomly from these notes for training/test sets

Data: gold standards

BRAT (Brat Rapid Annotation Tool) used to identify correct drug information

Input: clinical note

Expected output (for supervised learning): gold standard annotations
Data: gold standards

1. Develop annotation guidelines
   - When to highlight information
   - What defines different drug entities

2. Double annotation
   - 2 independent reviewers, evaluate annotation concordance

3. Revise guidelines if needed

4. Annotate training notes

5. Annotate test notes

Performance measures

Precision = \( \frac{\text{true positives}}{\text{true positives} + \text{false positives}} \)

Recall = \( \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \)

- Positive predictive value
- Sensitivity (true positive rate)
- Fraction of extracted output in gold standard
- Fraction of annotations that were correctly extracted
Performance measures

\[ F - \text{measure (F1)} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \]

- Quantify uncertainty with 95% bootstrap confidence intervals
  - Bootstrap notes (within drug)
  - Use 2.5 and 97.5 percentiles as interval bounds

Selecting tuning parameters

- Tune two parameters: window length and maximum edit distance
  - Create a grid of options for each parameter
  - Compute F-measure
  - Select parameters with best performance
- Maximize performance on training set
medExtractR: results

- Tacrolimus prescription patterns are much more simple than lamotrigine
- Allopurinol tested using tacrolimus tuning parameters

<table>
<thead>
<tr>
<th>medExtractR Extractions</th>
<th>DrugName</th>
<th>Strength</th>
<th>DoseAmount</th>
<th>Dose</th>
<th>Frequency</th>
<th>Intake/Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tacrolimus (training)</td>
<td>100</td>
<td>43</td>
<td>42</td>
<td>24</td>
<td>54</td>
<td>6</td>
</tr>
<tr>
<td>Lamotrigine (training)</td>
<td>77</td>
<td>27</td>
<td>39</td>
<td>53</td>
<td>44</td>
<td>32</td>
</tr>
<tr>
<td>Tacrolimus (test)</td>
<td>70</td>
<td>39</td>
<td>39</td>
<td>52</td>
<td>67</td>
<td>9</td>
</tr>
<tr>
<td>Lamotrigine (test)</td>
<td>77</td>
<td>34</td>
<td>44</td>
<td>52</td>
<td>42</td>
<td>29</td>
</tr>
<tr>
<td>Allopurinol (test)</td>
<td>187</td>
<td>79</td>
<td>84</td>
<td>35</td>
<td>106</td>
<td>12</td>
</tr>
</tbody>
</table>
MIMIC-III Clinical Care Database

- De-identified records corresponding to over 60,000 ICU stays
- Over 2 million clinical notes
- Institution: Beth Israel Deaconess Medical Center, Boston, MA
- Dataset available by request through MIT: https://mimic.physionet.org

Data

[patient_id]

taking drug

[note_id]

with [category]

patient_1

...

patient_n

note_1 (physician’s note)

note_2 (discharge summary)

...

note_n (physician’s note)

note_1 (physician’s note)

note_2 (nursing report)

...

note_m (discharge summary)

Categories have different likelihood of containing dose information
Note sampling procedure

3 drugs: tacrolimus, lamotrigine, oxcarbazepine

- **Tuning set**: 10 notes per drug
  - Randomly select notes one at a time
  - Manually review for presence of dosing information
  - If present, add to tuning set

- **Validation set**: 100 notes per drug
  - Randomly sample 50 discharge summaries
  - Randomly sample 50 from all other note categories

- Determine changes to annotation guidelines
- Annotate gold standards after tuning

MIMIC-III: Starting point

- Tuning set errors motivate next steps
- e.g. Tacrolimus 1 mg: One (1) capsule q daily

<table>
<thead>
<tr>
<th>Position</th>
<th>medExtractR</th>
<th>Gold Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:11</td>
<td>Tacrolimus</td>
<td>Tacrolimus</td>
</tr>
<tr>
<td>12:16</td>
<td>1 mg</td>
<td>1 mg</td>
</tr>
<tr>
<td>18:21</td>
<td>&lt;NA&gt;</td>
<td>One</td>
</tr>
<tr>
<td>23:24</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>34:41</td>
<td>&lt;NA&gt;</td>
<td>q daily</td>
</tr>
</tbody>
</table>

False negatives
Evaluation method

Present performance for different quantities for each drug:

1. No modification – “out of box” performance based on SD development

2. Tuning only
   - Smaller changes (dictionary updates, parameter selection)

3. Tuning plus customization
   - Adding or changing rules in the source code
   - Requires more advanced coding ability

4. Tuning plus customization
   - Text number followed by (digit) is a dose amount
MIMIC-III evaluation

- Performance with no tuning is not ideal
- F-makesures still above 0.80 benchmark

### Tacrolimus (n = 423 annotations)

<table>
<thead>
<tr>
<th></th>
<th>No tuning</th>
<th>Tuning only</th>
<th>Tuning plus customization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Precision</strong></td>
<td>.96 [.92, .99]</td>
<td>.93 [.89, .96]</td>
<td>.95 [.91, .98]</td>
</tr>
<tr>
<td><strong>Recall</strong></td>
<td>.77 [.71, .83]</td>
<td>.81 [.76, .85]</td>
<td>.89 [.84, .94]</td>
</tr>
<tr>
<td><strong>F-measure</strong></td>
<td>.85 [.81, .89]</td>
<td>.86 [.83, .90]</td>
<td>.92 [.88, .95]</td>
</tr>
</tbody>
</table>

### Lamotrigine (n = 381 annotations)

<table>
<thead>
<tr>
<th></th>
<th>No tuning</th>
<th>Tuning only</th>
<th>Tuning plus customization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Precision</strong></td>
<td>.87 [.82, .92]</td>
<td>.93 [.89, .97]</td>
<td>.94 [.90, .98]</td>
</tr>
<tr>
<td><strong>Recall</strong></td>
<td>.81 [.77, .85]</td>
<td>.83 [.78, .87]</td>
<td>.92 [.87, .96]</td>
</tr>
<tr>
<td><strong>F-measure</strong></td>
<td>.84 [.81, .87]</td>
<td>.88 [.84, .91]</td>
<td>.93 [.89, .96]</td>
</tr>
</tbody>
</table>

### Oxcarbazepine (n = 375 annotations)

<table>
<thead>
<tr>
<th></th>
<th>No tuning</th>
<th>Tuning only</th>
<th>Tuning plus customization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Precision</strong></td>
<td>.79 [.72, .86]</td>
<td>.97 [.94, .99]</td>
<td>.97 [.95, .99]</td>
</tr>
<tr>
<td><strong>Recall</strong></td>
<td>.83 [.79, .87]</td>
<td>.85 [.80, .89]</td>
<td>.92 [.88, .96]</td>
</tr>
<tr>
<td><strong>F-measure</strong></td>
<td>.81 [.76, .85]</td>
<td>.90 [.87, .93]</td>
<td>.95 [.92, .97]</td>
</tr>
</tbody>
</table>

- Some improvement with tuning alone
- Higher for lamotrigine and oxcarbazepine
- Largest improvement with tuning plus customization
Takeaways

- Without tuning, medExtractR performance is likely to be less than ideal, especially if building datasets for medication studies.
- Recommend at least performing tuning steps when using medExtractR for a new study. Customization is ideal, when possible.
- medExtractR approach provides a compromise between relying on “out-of-box” performance of existing medication extraction systems and having to manually create a validated dataset.

Contact

- Email: hannah.L.weeks@Vanderbilt.edu