Navigating Longitudinal Clinical Notes With An Automated Method For Detecting New Information

Rui Zhang\textsuperscript{a}, Serguei V. Pakhomov\textsuperscript{a,b}, Janet T. Lee\textsuperscript{c}, Genevieve B. Melton\textsuperscript{a,c}

\textsuperscript{a} Institute for Health Informatics, \\
\textsuperscript{b} College of Pharmacy, \textsuperscript{c} Department of Surgery, University of Minnesota, Minneapolis, MN, USA
Introduction

- Narratives in EHR systems provide clinicians with extensive patient information
- Clinicians can create redundant information by copying information from previous notes
- Redundant information may decrease healthcare efficiency
  - Increases the cognitive load on clinicians
  - Unidentified erroneous information could have adverse effects on patient safety

Introduction

- Abundant redundant information exists in clinical narrative
  - Inpatient notes (Wrenn, 2010)
  - Outpatient notes (Zhang, 2011&2012)
- Limited investigation into the source of redundant information
  - Can help understand clinicians’ behavior in generating notes
- A need for computational tools
  - Assist clinicians to synthesize complex patients
  - Navigate to notes with more new information

Objectives

- To understand “copy and paste” behaviors of clinicians
- To design an automated method
  - Navigate to notes with new information
  - Investigate new information patterns
Methods: system architecture

Notes

Annotation 1

Models

Text Pre-processing
- Sentence Splitter
- Note Format Handler
- Section Detector

Clinical expert manual annotated reference standard

Inter-rater reliability assessment

Performance Evaluation

Statistical N-gram language Methods
- Classic stopword removal
- TF-IDF stopword removal
- Lexical variation generation
- Heuristic rules

For a given patient m

The kth target note

Apply methods to identify new information in the target note

j previous longitudinal notes

Information navigation

New information pattern

Note: A1, A2, A3, ..., An

Note: B1, B2, B3, ..., Bn

Note: C1, C2, C3, ..., Cn

Note: N1, N2, N3, ..., Nn

...
Methods: data

- U Minnesota Medical Center Fairview Health Services (2005-10)
- Randomly selected 100 patients with angina, COPD (Chronic obstructive pulmonary disease), or diabetes for larger datasets
- Epic EHR system notes (~3000)
Methods: annotation 1

- New and clinically relevant information
- Based on their clinical judgment

Annotate one sample note

Annotate another 10 notes

Annotate 90 clinical notes overall

- Kappa: 0.80
- Percent agreement: 97%

40 for training and 50 for testing the language models
Methods: model to identify new information

For a given patient $m$:
- The $k$th target note
- $j$ previous longitudinal notes

Apply methods to identify new information in the target note:
- Classic stopword removal
- TF-IDF stopword removal
- Lexical variation generation
- Heuristic rules

Performance Evaluation


A bigram language model with classic stopword removal, TF-IDF stopword removal, application with lexical variation generation, and the adjustment through heuristic rules.

Accuracy: 0.83
Precision: 0.72
Recall: 0.71
F-Measure: 0.72
Methods: annotation 2

- Without seeing our results
- Two randomly selected patients

- Noted new information for 22-38th notes
- Used to compare results of information navigation

Noted new information for the 21st note

Noted any changes of new information for the 21st note
Methods: new information pattern analysis

Chronologically ordered notes of each patient (n₁, n₂, and nⱼ can be different numbers)

Patients

Build a language model

Predict new information

Calculate new information proportion (NIP)

Number of previous notes to build a model

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>...</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>A21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Arithmetic mean NIP scores

\[
NIP = \frac{\text{# sentences with new information}}{\text{# all sentences per note}}
\]

University of Minnesota
**Results:** clinicians’ copy & paste pattern

- 55% of information was from the most recent note.
- Additional 11% of information was from the previous 2-10 notes.
- Additional 4% was from the previous 11-20 notes.

The model is described by the equation:

\[ Y = -4.51\ln(x) + 44.6 \]

and has an R\(^2\) value of 0.98.
Results: NIP to navigate notes with new information

- Cyclical pattern
- High correlation with human judgment
- Source note of redundant information
Conclusions

• Clinicians copy information from the most recent notes
• New information in longitudinal notes had a logarithmic relationship with the length of historical notes
• NIP helps find notes with clinically relevant information and could be useful to navigate notes
• Further research
  - Develop more robust methods to detect new information
  - Classify types of new information (findings, medications, labs)
Acknowledgements

• Advisors
  ➢ Genevieve Melton-Meaux, MD, MA
  ➢ Serguei Pakhomov, PhD
• American Surgical Association Grant (Melton)
• U.S. NLM Grant (#R01 LM009623-01) (Pakhomov)
• UMN Doctoral Dissertation Fellowship
• Fairview Health Services/UMMC
• Contact Information
  ➢ Rui Zhang, MS
  ➢ zhan1386@umn.edu
Thank You!

Sculpture Garden @ Minneapolis, MN, USA