Patient Similarity-guided Decision Support

Tanveer Syeda-Mahmood, PhD
IBM Almaden Research Center
What is clinical decision support?

- Rule-based expert systems – curated by people, inferred by machines for CAD
- Practice guidelines – curated by people and presented by machines
- Rule-based alerts – curated by people, acted upon by machines
- Our approach – statistical decision support from EHR data combined with computable practice guidelines
  - A scalable way to leverage the knowledge in electronic health records
  - Personalized decision support
  - Developed in the domain of cardiology in consultations with cardiologists
  - Test deployed at the catheter lab at Kaiser Permanente’s San Francisco Medical Center
    - 3000 patients
  - Running in production at Cedars-Sinai Medical Center for the use case of similar patient cohort retrieval for clinical studies.
    - 1 million patients
  - Patient similarity technology in at least 2 IBM software products/solutions
Patient similarity for clinical decision support

Similar clinical data => similar patients => infer similarity in diagnosis, treatments and outcomes
Patient similarity algorithms

- Early fusion approaches using fully-supervised methods
  - Combine normalized features from all modalities into a single vector per sample (patient)
  - Given labeled pairs of similar feature vectors (by clinicians), compute pair-wise distance between samples.
  - Learn a metric in a projected space that ideally separates those vectors from the different similarity groups but groups those that are similar closer in this space
    - Metric learning can be based on many distance measures
      - Mahalanobis distance-based metric learning
      - Information-theoretic metric learning.

- Pros/Cons
  - Pros: Simple to model and use conventional machine learning framework
  - Cons: Cannot handle missing and spurious data well
    - Addition of new data means the metric has to be learned all over again - incremental update of learned distance metric difficult.
    - Computationally complexity and memory requirements can be excessive. Limitation of the size of the distance matrices that can be loaded.
    - Manually intensive as it requires clinicians to compare patients pairwise.
    - Not demonstrated for scalability so far.
Patient similarity algorithms

- Late fusion approaches (semi-supervised)
  - Compute similarity in each modality using separate distance metrics best suited for their respective domain data.
    - Produce separate ranked list of similar patients per modality
    - Fuse ranked lists from modalities to do an overall ranking of the patients
    - Weak supervision by labeling the data based on diagnosis. Use diagnosis as a key element of patient similarity.
      - The diagnosis are either in structured records or can be extracted from reports.

- Pros/Cons
  - Pros: Works even if not all modalities are present for a given patient.
    - Allows for customization of the metric as per modality and clinical content.
      - What makes two EKGs similar is not the same as two echo videos similar.
    - Scalable way to collect ground truth from reports and structured data and doesn’t need pairwise similarity comparisons to be made by clinicians.
  - Cons: Doesn’t handle time-sensitive aspects.
    - Similarity in modalities have to be over the same period of time in order to be mutually reinforcing
      - Patients who have similar EKGs and similar echocardiograms over a consistent period in time are more similar than those that match at different points in time.
EKG similarity

- **EKG similarity search**
  - Invariant to heart rate and signal strength. Computes it automatically
  - Captures perceptual similarity of shapes.
  - No extraction of PR, QT and other intervals needed
  - No built-in rules for recognition of patterns.
  - Uses non-rigid shape matching.

Matching patients retrieved based on EKG and their diagnosis

| Rank 1-ID: 426 (BBB) |
| Rank 2-ID: 404 (BBB) |
| Rank 3-ID: 364 (BBB) |
| Rank 4-ID: 508 (BBB) |
| Rank 5-ID: 191 (MI, AF) |
| Rank 6-ID: 372 (MI) |

BBB:67% MI: 33% AF:16% in EMBC’07
Modality-specific similarity: Text similarity

- Extract concepts from textual reports
  - Large vocabulary-driven extraction of diseases, drugs, symptoms, family history, measurements
  - Vocabularies formed from SNOMED CT, LOINC, RxNorm, ICD9, findings from mining millions of reports
  - Negation and family history references filtered.
- Find similar patients using textual similarity based on cosine distance.

Extracted diseases and severity (positive evidence)
mitral regurgitation, moderate pulmonary hypertension, mitral stenosis

Extracted diseases (negative evidence)
pulmonic valvular insufficiency, pericardial effusion

Extracted measurements
RVID<3.0cm, Age=71, gender=F, etc.

In EMBC ‘2010
Doppler pattern similarity

CVPR '2010

First work to do automatic valvular disease recognition from Doppler
Patient similarity algorithms

- **Multimodal time-sensitive fusion**
  - Compute similarity in each modality using separate distance metrics best suited for their respective domain data.
    - Produce separate ranked list of similar patients per modality
  - Model pairwise similarity between patients as a **time-varying similarity function**.
  - Fuse ranked lists from modalities to do an overall ranking of the patients using time-sensitive multimodal fusion
  - Weak supervision by labeling the data based on diagnosis. Use diagnosis as a key element of patient similarity.
    - The diagnosis are either in structured records or can be extracted from reports.

- **Pros/Cons**
  - Pros: Works even if not all modalities are present for a given patient.
    - Allows for customization of the metric as per modality and clinical content.
      - What makes two EKGs similar is not the same as two echo videos similar.
    - Scalable way to collect ground truth from reports and structured data and doesn’t need pairwise similarity comparisons to be made by clinicians.
  - Cons: Can be difficult to explain.
Multimodal time-sensitive fusion: Overall approach

- **Top K list from EKG search**
- **Top K from echo search**
- **Top K list from report search**
- **Top K from diagnosis search**

Map to patient clinical timeline

Multimodal fusion

- **Top K ranked Patient time periods**
- **Top N diseases**
- **Top N drugs**

Collaborative filtering

EKGs query

Echos query

Reports query

Diagnosis query

Patient Clinical History models
Time-varying modeling in AALIM
Fusing similarity by multi-modal time-sensitive fusion

Each modality is a CER variable.
Representing time-varying CER variables for similarity fusion

CER variables: $V_1, V_2, \ldots, V_M$ the set of values taken by a variable $V_i \rightarrow v_{i1}, v_{i2}, \ldots, v_{ik_i}$

diseases, drugs

A longitudinal patient model (LPM) for a patient P can be denoted by a set of unit time series

$$LPM(P) = \{S_P(v_{ik}, t) \mid S_P(v_{ik}, t) > 0 \text{ for some } t, T_{\text{min}}(P) \leq t \leq T_{\text{max}}(P)\}$$

$$S_P(v_{ik}, t) = \begin{cases} 
1 & \text{if a value of } v_{ik} \text{ is recorded for CER variable } V_i \text{ for patient } P \text{ at time } t \\
0 & \text{otherwise}
\end{cases}$$

Eg. a diagnosis of cardiomyopathy over a period of time

$$S_P(v_{\text{disease}}, \text{cardiomyopathy}, t) = 1, t_1 \leq t \leq t_2$$
**Multimodal fusion algorithm**

Suppose a query patient $Q$ has

$$N_Q = |\bar{V}_Q| = \{V_i \mid \exists_{v_{ik}} S_Q(v_{ik}, t) > 0 \text{ for some } t\}$$

the number of different values exhibited per CER variable in a patient $Q$

$$n_Q(V_i) = |\bar{v}_Q| = \{v_{ik} \mid S_Q(v_{ik}, t) > 0 \text{ for some } t\}$$

For example, a patient diagnosed with mitral stenosis, hypertension, and aortic stenosis and on medications: warfarin, furosemide, atenolol and amoxicillin during a 10 year period would have

$$N_Q = 2 \quad n_Q(V_{\text{disease}}) = 3 \quad n_Q(V_{\text{drugs}}) = 4$$

Similarity in the values of a CER variable $V_i$

exhibited by a candidate patient $P$ and a specific query patient $Q$ can be denoted by $d_{PQ}(v_{ik}, v_{il})$

For each candidate patient $P$ and CER variable $V_i$

we identify all time points in the patient’s time span where a match to one of the values of the query patient $Q$ for variable $V_i$ exists
Multimodal fusion algorithm

For each value $v_{ik}$ of the variable $V_i$ in the set $\vec{V}_{Qi}$ possessed by the query patient Q, we can record all time periods in the timeline of a candidate patient P where there is a match to this value as a function $m_{PQ}(v_{ik}, t)$

$$m_{PQ}(v_{ik}, t) = \begin{cases} d_{PQ}(v_{ik}, v_{il}) & \text{if } 1 \geq d_{PQ}(v_{ik}, v_{il}) \geq \rho \text{ and } S_P(v_{il}, t) = 1 \\ 0 & \text{otherwise} \end{cases}$$

Then the extent of match of a patient P to a query patient Q based on the CER variable $V_i$

$$C_{PQ}(V_i, t) = \sum_{v_{ik}} m_{PQ}(v_{ik}, t)$$

For all CER variables $\vec{V}_{Qi}$ exhibited by Q as

$$U_{PQ}(t) = \frac{\sum_{i} C_{PQ}(V_i, t)}{N_Q}$$

We can now form a patient cohort

$$H(Q) = \{ P \mid U_{PQ}(t) > \lambda \text{ for some } t \}$$
Using similarity by multimodal fusion for clinical decision support

Query patient had:
Hyperlipidemia
Hypertension
Atherosclerosis
Results

- Patient data set:
  - 1996 patients
  - 12 channel EKG time series: 25,016
  - Echo-cardiographic sequences: 5346
  - CW Doppler images: 34,540
  - Textual reports: 100,042

- Evaluation of Patient similarity by multimodal fusion:
  - Automatic: Comparing against ground truth data about the patients’ diseases.
  - Manual: Validation of the top K lists by clinicians.
    - Clinician examines clinical record of the patient
    - Clinician annotates the top K lists returned by the fusion algorithm as

<table>
<thead>
<tr>
<th>F</th>
<th>factual (present for this person)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NF</td>
<td>non factual</td>
</tr>
<tr>
<td>R</td>
<td>Relevant (to consider with similar cardiac patients)</td>
</tr>
<tr>
<td>NR</td>
<td>Non-relevant</td>
</tr>
<tr>
<td>U</td>
<td>Unlikely, unrelated to patient’s cardiac condition</td>
</tr>
</tbody>
</table>
Evaluation of top K disease lists returned by patient similarity

- Given top K matches to a query Q with disease label: \( L_Q = (l_{q1}, l_{q2}, \ldots, l_{qN}) \)
- Let the distribution of disease labels among the top K matches be: \( L_M = (l_1, l_2, \ldots, l_M) \)

- Metrics used:
  - Recall: Fraction of overlap of the query disease labels with those of the matches
    \[
    \text{Recall} = \frac{|L_Q \cap L_M|}{|L_Q|}
    \]
  - New predictions
    \[
    \text{Discoverability} = \frac{|L_M - L_Q|}{|L_M|}
    \]
  - Valid predictions:
    \[
    \text{Validity} = \sum_{l_i \in L_M - L_Q} \max_{l_qi \in L_Q} (p(l_i, l_qi)) \frac{|L_M - L_Q|}{|L_M - L_Q|}, |L_M - L_Q| > 0
    \]
Evaluation Studies per modality: EKG

Dataset: 25,990 ECGs

Set of disease labels in the returned matches \( L_M = (l_1, l_2, \ldots, l_M) \)

Set of disease labels in query: \( L_Q = (l_{q1}, l_{q2}, \ldots, l_{qN}) \)

Accuracy: \( \text{Recall} = \frac{|L_Q \cap L_M|}{|L_Q|} \)

Potential co-morbidity discoveries:

\[ \text{discoverability} = \frac{|L_M - L_Q|}{|L_M|} \]

Efficacy: \( \text{discoverability} \times \text{likelihood of the co-occurrence of any disease pairs} \)

\[ \text{validity} = \sum_{l_i \in L_M - L_Q \atop l_{qj} \in L_Q} \max(p(l_i, l_{qj})) \]

85% of them were completely consistent with the official human interpretation (100% recall), and nearly all EKGs had at least 50% overlap in their match set with their own disease labels.
Evaluation studies on clinical decision support by multimodal fusion

<table>
<thead>
<tr>
<th>Legend: F: factual (present for this person)</th>
<th>NF: non-factual</th>
</tr>
</thead>
<tbody>
<tr>
<td>F: Relevant (to consider with similar cardiac patients)</td>
<td>NR: non-relevant</td>
</tr>
<tr>
<td>U: Unlikely, unrelated to patient's cardiac condition</td>
<td></td>
</tr>
</tbody>
</table>

### Clinician annotation of top K lists

<table>
<thead>
<tr>
<th>Condition</th>
<th>F/NF/R/NR/U</th>
<th>Condition</th>
<th>F/NF/R/NR/U</th>
</tr>
</thead>
<tbody>
<tr>
<td>272.4 Other and unspecified hyperlipidemia (Disorders of lipid metabolism)</td>
<td>F</td>
<td>NF</td>
<td>R</td>
</tr>
<tr>
<td>414.01 Of native coronary artery (Other forms of chronic ischemic heart disease; Coronary atherosclerosis)</td>
<td>F</td>
<td>NF</td>
<td>R</td>
</tr>
<tr>
<td>424.1 Aortic valve disorders (Other diseases of endocardium)</td>
<td>F</td>
<td>NF</td>
<td>R</td>
</tr>
<tr>
<td>782.3 Edema (Symptoms involving skin and other integumentary tissue)</td>
<td>F</td>
<td>NF</td>
<td>R</td>
</tr>
<tr>
<td>428.0 Congestive heart failure, unspecified (Heart failure)</td>
<td>F</td>
<td>NF</td>
<td>R</td>
</tr>
<tr>
<td>427.31 Atrial fibrillation (Cardiac dysrhythmias; Atrial fibrillation and flutter)</td>
<td>F</td>
<td>NF</td>
<td>R</td>
</tr>
<tr>
<td>308.9 Unspecified acute reaction to stress (Acute reaction to stress)</td>
<td>F</td>
<td>NF</td>
<td>R</td>
</tr>
<tr>
<td>414.01 Of unspecified type of vessel, native or graft (Other forms of chronic ischemic heart disease; Coronary atherosclerosis)</td>
<td>F</td>
<td>NF</td>
<td>R</td>
</tr>
<tr>
<td>780.4 Dizziness and giddiness (General symptoms)</td>
<td>F</td>
<td>NF</td>
<td>R</td>
</tr>
<tr>
<td>530.81 Esophageal reflux (Diseases of esophagus; Other specified digestive system disease)</td>
<td>F</td>
<td>NF</td>
<td>R</td>
</tr>
<tr>
<td>443.9 Peripheral vascular disease, unspecified (Other peripheral vascular disease)</td>
<td>F</td>
<td>NF</td>
<td>R</td>
</tr>
<tr>
<td>424.0 Mitral valve disorders (Other diseases of endocardium)</td>
<td>F</td>
<td>NF</td>
<td>R</td>
</tr>
<tr>
<td>593.9 Unspecified disorder of kidney and ureter (Other disorders of urinary system)</td>
<td>F</td>
<td>NF</td>
<td>R</td>
</tr>
<tr>
<td>465.0 Acute bronchitis (Acute bronchitis and bronchiolitis)</td>
<td>F</td>
<td>NF</td>
<td>R</td>
</tr>
<tr>
<td>427.9 Cardiac dysrhythmia, unspecified (Cardiac dysrhythmias)</td>
<td>F</td>
<td>NF</td>
<td>R</td>
</tr>
<tr>
<td>784.0 Headache (Symptoms involving head and neck)</td>
<td>F</td>
<td>NF</td>
<td>R</td>
</tr>
<tr>
<td>412.0 Old myocardial infarction</td>
<td>F</td>
<td>NF</td>
<td>R</td>
</tr>
<tr>
<td>780.2 Syncope and collapse (General symptoms)</td>
<td>F</td>
<td>NF</td>
<td>R</td>
</tr>
<tr>
<td>429.9 Heart disease, unspecified (ill-defined descriptions and complications)</td>
<td>F</td>
<td>NF</td>
<td>R</td>
</tr>
</tbody>
</table>
Results of manual validation by clinicians of top K lists from multimodal fusion

<table>
<thead>
<tr>
<th>Patient</th>
<th>Relevant</th>
<th>Non-Relevant</th>
<th>Unlikely/Unrelated</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Factual</td>
<td>Discovered</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>0</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>1</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>3</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>Average</td>
<td>50%</td>
<td>22%</td>
<td>72%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Summary

- Multi-modal time-sensitive fusion combines similarity of clinical data in multiple modalities while respecting time overlap in their occurrence.

- Patient similarity is a scalable way to achieve clinical decision support without any built-in rules.

- It is personalized for the patient-specific conditions.

- An example of how knowledge from electronic health records could be leveraged for meaningful use.