Time Series Analysis of Nursing Notes for Mortality Prediction via a State Transition Topic Model

Yohan Jo, Natasha Loghmanpour, Carolyn P. Rose
Problem

• Mortality Prediction for Patients in Intensive Care Units
  – At a given time, predict whether a patient would die within a day, a month, a year, or would survive

• In the Field
  – Nurses depend on alarms from bedside monitors
    • Bedside monitors make a lot of false alarms
    • Making nurses desensitized to important alarms
  – Nurses are busy
    • Nurses are a few and shift often
    • Making it hard to trace each patient’s condition
  – Nurses need to prioritize their work
Previous Studies on Automatic Mortality Prediction

- **Waveform Data**
  - **Blood pressure**: Lehman, L.-W. et al. (2014)

  Time series analysis proves promising

- **Textual Data**
  - **Nursing notes**: Lehman, L. et al. (2012)

  No report on time series analysis
Nursing Notes as a Prediction Resource

**PROB:** S/P AVR
FUNCTIONAL HEALTH PATTERNS AND HISTORY COMPLETED IN CHART.

**NEURO:** PT AWAKE, FOLLOWING COMMANDS. MAE, NODS APPROPRIATELY. PERL.

**CV:** SR NO VEA NOTED. CONT ON MILRINONE AND NEO. CO/CI [**7–26**]. K REPLACED MULTIPLE TIMES. CT DRAINING S/S DRAINAGE. PACER OFF. MORPHINE FOR PAIN X3 WITH GOOD EFFECT.

**RESP:** PT WEANING. LUNGS WITH WHEEZES, CLEARED WITH COMBIVENT INHALER. SUCTION FOR THICK CLEAR/YELLOW SPUTUM. PT STILL ACIDOTIC, IMV 16, TV 700.

**GI:** NGT TO LOW CONT SUCTION, NO DRAINAGE,

**ENDO:** INITIAL BS ELEVATED AND TREATED PER PROTOCOL.

**GU:** ADEQUATE AMOUNT OF CLEAR YELLOW URINE. PLACEMENT CHECKED–GOOD.

**SOCIAL:** FAMILY HERE TO VISIT.

**ASSESSMENT:** WEANING SLOWLY

**PLAN:** RECHECK ABGS, LYTEs.
SUCTION PRN.
MED FOR PAIN.
CONT VENT WEAN.

-- MIMIC2 Dataset by Physionet

- Family support
- Mental fitness
- Facial expressions
- Nurses’ intuitions & plans
Previous Study (Ghassemi et al., 2014)

No temporal aspect is considered
This Study

As the foundation of the LTI identity, our logo serves as the most concise visual expression of our brand. Flexible, reliable and creative, the logo is an essential element for any brand communication.
State Transition Topic Model (STTM)

- Generative Process
  - Each state represents a patient’s (latent) condition and has a topic distribution
  - At each time point, a patient enters into a state according to the transition probabilities of the previous state (HMM)
  - In the new state, nursing notes are generated from the topic distribution (LDA)
State Transition Topic Model (STTM)

We can learn
- $\phi_j$: word distribution of each topic
- $\theta_c$: topic distribution of each state
- $\theta_{t,z}^m$: topic distribution of each document
- $\pi_c$: state transition probability distribution
- State of each document
  (⇒ used as a feature for classification)

- **How different from LDA + clustering?** Topics learned by considering state transitions may have unique characteristics.
- **How different from other temporal topic models?** A document’s topics do not directly determine the next document’s topics.
- Given observed documents, we may find meaningful state representation and trend of state transitions.
Task 1.

TEMPORAL INFORMATION EXTRACTION BY STTM
Data

- **MIMIC II Clinical Database**
  - Clinical data of ICU patients collected between 2001 and 2008

<table>
<thead>
<tr>
<th></th>
<th>1 day</th>
<th>1 week</th>
<th>1 month</th>
<th>6 months</th>
<th>1 year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Died within</td>
<td>%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>%</td>
<td>3</td>
<td>8</td>
<td>14</td>
<td>19</td>
<td>22</td>
</tr>
</tbody>
</table>

- # of sequences (\# of patients): 8,808
- Avg length of sequences: 11 (Stddev=23)
- Avg length of each document: 1,548 chars (Stddev=904)

One time point is 12 hrs.
State-Aware Topics

<table>
<thead>
<tr>
<th>T0. Infant admission</th>
</tr>
</thead>
<tbody>
<tr>
<td>infant, nicu, delivery, cbc, normal, sepsis, maternal, born, admission, risk, baby, gbs, PHI_DATE, apgars, pregnancy</td>
</tr>
</tbody>
</table>

Admit/transfer note Infant admitted from l&D for sepsis eval. [**Name8 (MD) 63**] MD/NNP note for hx and physical.

<table>
<thead>
<tr>
<th>T1. Stability</th>
</tr>
</thead>
<tbody>
<tr>
<td>feeds, cares, infant, well, voiding, continue, parents, feeding, active, stable, support, cc, ra, far, crib, swaddled, mom, fen</td>
</tr>
</tbody>
</table>

continue to encourage po feeds. 4: G/D temps stable in an open crib. alert and active with cares. sleeps well inbetween.

<table>
<thead>
<tr>
<th>T2. Family visiting</th>
</tr>
</thead>
<tbody>
<tr>
<td>family, patient, but, today, pts, sats, team, placed, iv, mg, bp, mask, off, afternoon, night, started, status, able, day, continues</td>
</tr>
</tbody>
</table>

nnp Pt expired at 7:40pm. Family at bedside with pt. Emotional support given to family at this very difficult time.

<table>
<thead>
<tr>
<th>T3. Admission</th>
</tr>
</thead>
<tbody>
<tr>
<td>ccu, micu, PHI_DATE, cath, s/p, c/o, PHI_HOSPITAL, admitted, admission, sob, floor, arrival, hct, arrived, ns, wife</td>
</tr>
</tbody>
</table>

Pt transferred to [**Hospital1 22**] for ERCP which was done on arrival to MICU.

<table>
<thead>
<tr>
<th>T4. Newborn jaundice</th>
</tr>
</thead>
<tbody>
<tr>
<td>bili, infant, cpap, baby, isoulette, caffeine, servo, nested, phototherapy, enteral, mom, feeds, cc/kg/day, cc/k/d, will</td>
</tr>
</tbody>
</table>

#6-O: under single phototherapy, bili pending this am. sl jaundiced. #4-O; temp stable in servo isoulette, active and alert

<table>
<thead>
<tr>
<th>T5. Social activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>he, she, that, her, his, have, been, but, had, very, also, will, PHI_LASTNAME, if, need, would, after, about, did, they, it</td>
</tr>
</tbody>
</table>

NPN Pt left AMA, he felt that he needed to get home to help take care of his daughters.

<table>
<thead>
<tr>
<th>T6. Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>am, social, place, blood, progress, due, care, up, time, pm, fluid, npn, per, noted, cxr, id, area, cc, pain, urine, site, currently</td>
</tr>
</tbody>
</table>

NURSING PROGRESS NOTE 11 PM - 7 AM NO COMPLAINTS HR 80'S...SBP 90-100'S..

<table>
<thead>
<tr>
<th>T7. Respiratory support</th>
</tr>
</thead>
<tbody>
<tr>
<td>thick, vent, secretions, peep, coarse, sputum, abg, suctioned, yellow, goal, tube, trach, remains, gtt, eyes, ams, skin</td>
</tr>
</tbody>
</table>

Resp Care: Pt continues trached and on ventilatory support with simv 600x14/+5 peep/5 psv/fio2 .4 with acceptable abg

<table>
<thead>
<tr>
<th>T8. Sepsis</th>
</tr>
</thead>
<tbody>
<tr>
<td>npo, settings, fio2, pn, infusing, secretions, cbg, rate, vent, amp, coarse, respiratory, gas, lytes, received, ett, white, simv, cloudy</td>
</tr>
</tbody>
</table>

Respiratory Care Note Pt continues on SIMV 22/5 R 34 and FIO2 33-44%. BS are coarse. Pt. sx’d for mod.

<table>
<thead>
<tr>
<th>T9. Laboratory tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>pain, sbp, gtt, neuro, clear, cough, intact, oob, mg, pulses, Foley, neo, bs, effect, ct, bases, given, activity, mae, lopressor, po</td>
</tr>
</tbody>
</table>

7p-7a Neuro: a+ox3. Pain controlled with Percocet. CV: sbp initially 140's, 2.5mg iv with effect. SR with long PR, rate 80's.
State Transitions Learned from Nursing Notes

Figure 4 is the state transition graph constructed by the State-Tags model. The transitions are learned from nursing notes and show the progression of a patient's conditions. The nodes represent different topics, and the arrows indicate the flow of information and changes in the patient's condition.
OUR LOGOS

As the foundation of the LTI identity, our logo serves as the most concise visual expression of our brand. Flexible, reliable and creative, the logo is an essential element for any brand communication.

STTM vs. LDA

<table>
<thead>
<tr>
<th>Infant admission</th>
<th>Admission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Admission</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Family visiting</th>
<th>Social activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social activities</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Newborn jaundice</th>
<th>Sepsis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newborn jaundice</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stability</th>
<th>Stability1</th>
<th>Stability2</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Laboratory tests</th>
<th>Laboratory tests1</th>
<th>Laboratory tests2</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Summary</th>
<th>Summary</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Respiratory support</th>
<th>Sounds in organs</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Summary</th>
<th>Respiratory support</th>
</tr>
</thead>
</table>

Carnegie Mellon University
Language Technologies Institute
State Transitions Learned from Nursing Notes

Admission

Lab tests

Resp support

Family visiting

S2. Lab tests
Resp support

S7. Sepsis
Newborn jaundice

S3. Resp support

S4. Newborn jaundice
Sepsis

S5. Infant admission

S6. Stability
Newborn jaundice

S8. Stability

Infant admission
Yet, the labels should not be overemphasized as if each state distribution, node labels reflect the relative proportions of topics. As the foundation of the LTI identity, our logo serves as the most concise visual expression of our brand. Flexible, reliable and creative, the logo is an essential element for any brand communication. As the foundation of the LTI identity, our logo serves as the most concise visual expression of our brand. Flexible, reliable and creative, the logo is an essential element for any brand communication.
State Transitions Learned from Nursing Notes

**Figure 4** is the state transition graph constructed by the model to illustrate the transitions and topics learned from the nursing notes. The states and their transitions are visualized in the diagram, with labels indicating various conditions such as admission, infant admission, stability, sepsis, lab tests, and newborn jaundice. The diagram helps in understanding the communication and support changes in different stages of a patient's journey.
As the foundation of the LTI identity, our logo serves as the most concise visual expression of our brand. Flexible, reliable and creative, the logo is an essential element for any brand communication.
State Transitions Learned from Nursing Notes

S0. Resp support
S1. Newborn jaundice
S2. Lab tests
S3. Resp support
S4. Newborn jaundice Sepsis
S5. Infant admission
S6. Stability
S7. Sepsis
S8. Stability
S9. Admission

Lab tests
Resp support
Family visiting
Newborn jaundice
Newborn jaundice Sepsis
Newborn jaundice
Stability
Stability
Sepsis

Figure 4 is the state transition graph constructed by the model. This visual representation helps illustrate the transitions between different states, which are likely to correspond to changes in a patient's condition or progress through their hospital stay. Each state (node) represents a specific scenario or condition, and the arrows indicate the probabilities of moving from one state to another.
State Transitions Learned from Nursing Notes

S0. Resp support
Family visiting

S1. Newborn jaundice
Stability

S2. Lab tests
Resp support

S3. Resp support

S4. Newborn jaundice
Sepsis

S5. Infant admission

S6. Stability
Newborn jaundice

S7. Sepsis
Newborn jaundice

S8. Stability

S9. Admission
Lab tests

Laboratory tests

Figure 4 is the state transition graph constructed by the LTI algorithm, which learns state transitions from nursing notes (adapted from Dagan et al., 2013). As the foundation of the LTI identity, our logo serves as the most concise visual expression for the state transitions.
Task 2.

MORTALITY PREDICTION WITH TEMPORAL INFORMATION
Features

- **GT**
  - Standard Topics (50 topics by LDA) – Baseline

- **GT+ST**
  - Standard Topics + State Transitions (10 topics and 10 states by STTM)

- **GT+SA**
  - Standard Topics + State Transitions + State-Aware Topics (10 topics and 10 states by STTM)

- **GT+SA+**
  - Same as GT+SA but STTM is trained only on long sequences
Mortality Prediction with Combined Features

<table>
<thead>
<tr>
<th></th>
<th>$GT$</th>
<th>$GT+ST$</th>
<th>$GT+SA$</th>
<th>$GT+SA+$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-Day</td>
<td><strong>0.7325</strong></td>
<td>0.7218</td>
<td>0.7224</td>
<td>0.7235</td>
</tr>
<tr>
<td>1-Week</td>
<td>0.7781</td>
<td>0.7772</td>
<td><strong>0.7811</strong></td>
<td>0.7784</td>
</tr>
<tr>
<td>1-Month</td>
<td>0.7820</td>
<td><em>0.7860</em></td>
<td><em>0.7866</em></td>
<td><strong>0.7871</strong></td>
</tr>
<tr>
<td>6-Month</td>
<td>0.7882</td>
<td>0.7884</td>
<td><strong>0.7921</strong></td>
<td><em>0.7912</em></td>
</tr>
<tr>
<td>1-Year</td>
<td>0.7905</td>
<td>0.7912</td>
<td>0.7930</td>
<td><strong>0.7939</strong></td>
</tr>
</tbody>
</table>

- Metric: AUC (Area Under ROC Curve)

$$AUC = \frac{\sum_{i=1}^{n^+} \sum_{j=1}^{n^-} 1(f(x_i) > f(x_j))}{n^+n^-}$$

- GT: Standard Topics (50 topics by LDA) – Baseline
- GT+ST: Standard Topics + State Transitions (10 topics and 10 states by STTM)
- GT+SA: Standard Topics + State Transitions + State-Aware Topics (10 topics and 10 states by STTM)
- GT+SA+: Same as GT+SA but STTM is trained on longer sequences
Mortality Prediction by Terms and Time Points

Short-Term

Long-Term

1-Day Mortality Prediction

1-Month Mortality Prediction

1-Week Mortality Prediction

6-Month Mortality Prediction

1-Year Mortality Prediction

<table>
<thead>
<tr>
<th>Time Point</th>
<th>GT</th>
<th>GT+SA</th>
<th>GT+SA+</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13-14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17-18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19-20</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AUC values:

- **1-Day**: GT (0.80), GT+SA (0.84), GT+SA+ (0.79)
- **1-Week**: GT (0.78), GT+SA (0.81), GT+SA+ (0.79)
- **1-Month**: GT (0.74), GT+SA (0.79), GT+SA+ (0.78)
- **6-Month**: GT (0.72), GT+SA (0.79), GT+SA+ (0.78)
- **1-Year**: GT (0.74), GT+SA (0.81), GT+SA+ (0.80)

Significant differences confirmed for GT and GT+SA+.
Task 3.

MORTALITY PREDICTION BY INDIVIDUAL FEATURES
Mortality Prediction with Individual Features

<table>
<thead>
<tr>
<th></th>
<th>n-grams</th>
<th>10 Topics</th>
<th>50 Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>StandTopic</td>
<td>StateTopic</td>
</tr>
<tr>
<td>1-Day</td>
<td>0.563</td>
<td>0.709</td>
<td><strong>0.711</strong></td>
</tr>
<tr>
<td>1-Week</td>
<td>0.651</td>
<td>0.746</td>
<td>0.749</td>
</tr>
<tr>
<td>1-Month</td>
<td>0.691</td>
<td>0.753</td>
<td>0.759</td>
</tr>
<tr>
<td>6-Month</td>
<td>0.726</td>
<td>0.766</td>
<td>0.767</td>
</tr>
<tr>
<td>1-Year</td>
<td>0.732</td>
<td>0.772</td>
<td>0.771</td>
</tr>
</tbody>
</table>

- **n-grams**: N-grams
- **StandTopic**: Standard Topics (by LDA)
- **StateTopic**: State-Aware Topics (by STTM with 10 states)
- **StateTrans**: State Transitions (by STTM with 10 states)
- **StateAll**: State-Aware Topics + State Transitions
Enrichment

(a) Enrichment of state-aware topics

(b) Enrichment of states

- **Enrichment of feature** $w$

  \[
  \frac{\text{proportion of feature } w \text{ for patients who died}}{\text{total proportion of feature } w}
  \]
Conclusion

- STTM tends to combine/split topics appearing in similar/different trajectories.
- STTM reveals a meaningful trend of patients’ states and state transitions latent in nursing notes.
- The learned temporal information is beneficial for long-term mortality prediction, but not much in short-term prediction.
- The learned states indeed have different levels of correlations with mortality.
- STTM can be applied to any data stream.
Limitations

• No improvement in mortality prediction when the number of topics is increased from 10 to 50
  – More evaluation is needed on different data and in different aspects in order to better understand the scalability of STTM.

• No improvement when applied to NICUs and the others separately
  – This might be due to the reduced data size and the sparsity of the feature space (e.g., 100 possible state transitions).
  – More sophisticated approaches are desirable to use temporal information for different ICU types.
In the paper

- Comparisons among different temporal topic models
- Detailed analysis of individual textual features
References

(a) Statistics of retrieved nursing notes

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td># of sequences (=# of patients)</td>
<td>8,808</td>
</tr>
<tr>
<td># of notes</td>
<td>187,808</td>
</tr>
<tr>
<td># of merged documents</td>
<td>97,769</td>
</tr>
<tr>
<td>Avg length of merged documents</td>
<td>1,548 chars (Stddev=904)</td>
</tr>
<tr>
<td>Avg length of sequences</td>
<td>11 (Stddev=23)</td>
</tr>
</tbody>
</table>

(b) Characteristics of ICU types

<table>
<thead>
<tr>
<th></th>
<th>NICU</th>
<th>CSRU</th>
<th>MICU</th>
<th>CCU</th>
<th>FICU</th>
<th>SICU</th>
</tr>
</thead>
<tbody>
<tr>
<td># of patients</td>
<td>2332</td>
<td>2244</td>
<td>1918</td>
<td>1358</td>
<td>711</td>
<td>245</td>
</tr>
<tr>
<td>Avg. stay</td>
<td>10</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

(c) Percentages of patients who died within certain periods of time after admission

<table>
<thead>
<tr>
<th>Died within</th>
<th>1 day</th>
<th>1 week</th>
<th>1 month</th>
<th>6 months</th>
<th>1 year</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>0.03</td>
<td>0.08</td>
<td>0.14</td>
<td>0.19</td>
<td>0.22</td>
</tr>
</tbody>
</table>