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

Yohan Jo, Natasha Loghmanpour, Carolyn P. Rose

Carnegie Mellon University
Language Technologies Institute

Sponsored by
Naval Research Laboratory
Washington, DC

Problem

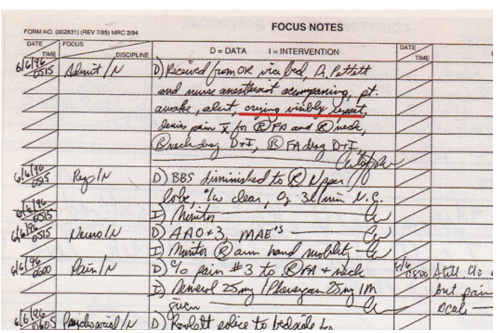
- Mortality Prediction for Patients in Intensive Care Units
  - At a given time, predict whether a patient would die within a day, a month, a year, or would survive

- 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
  - Heart rates: Chia, C.-C., & Syed, Z. (2011), Chia, C. et al. (2014), Syed, Z. et al. (2009), Syed, Z., & Guttag, J. V. (2010), Syed, Z. et al. (2011)
  - Blood pressure: Lehman, L.-W. et al. (2014)

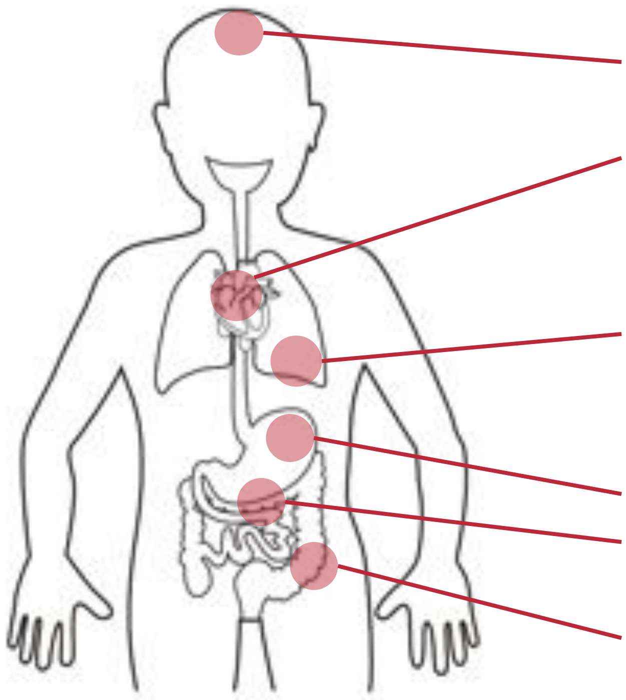
  Time series analysis proves promising



- Textual Data
  - Nursing notes: Lehman, L. et al. (2012)
  - Radiology report: Ghassemi, M. et al. (2014)

  No report on time series analysis

Nursing Notes as a Prediction Resource

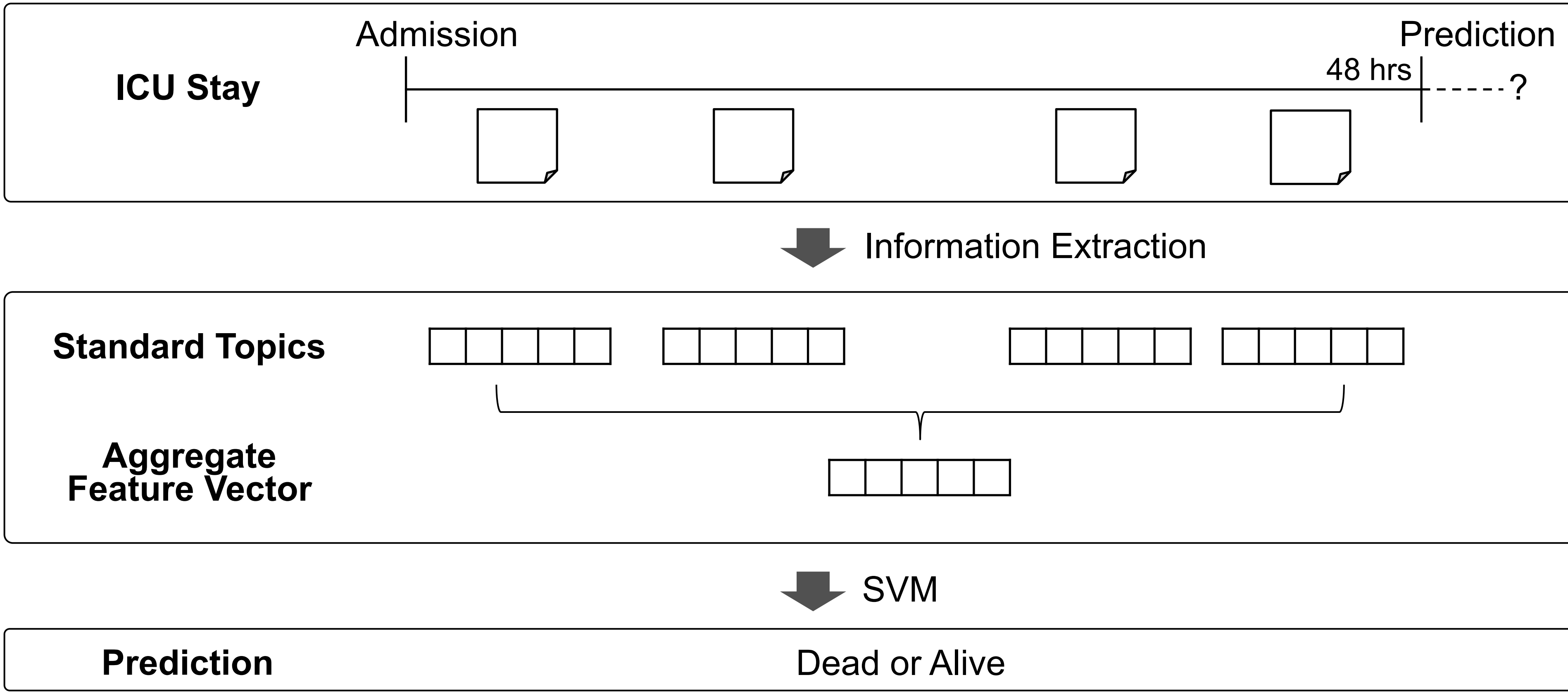


PROB: S/P AVR
FUNCTIONAL HEALTH PATTERNS AND HISTORY COMPLETED IN CHART.
NEURO: PT AWAKE, FOLLOWING COMMANDS. MAE, NODS APPROPRIATELY. PERL.
CV: SR NO VEA NOTED. CONT ON MILRINONE AND NEO. CO/CI [**7-26**]. K REPLACED MULTIPLE TIMES. CT DRAINING S/S DRAINAGE. PACER OFF. MORPHINE FOR PAIN X3 WITH GOOD 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.

- Family support
- Mental fitness
- Facial expressions
- Nurses' intuitions & plans

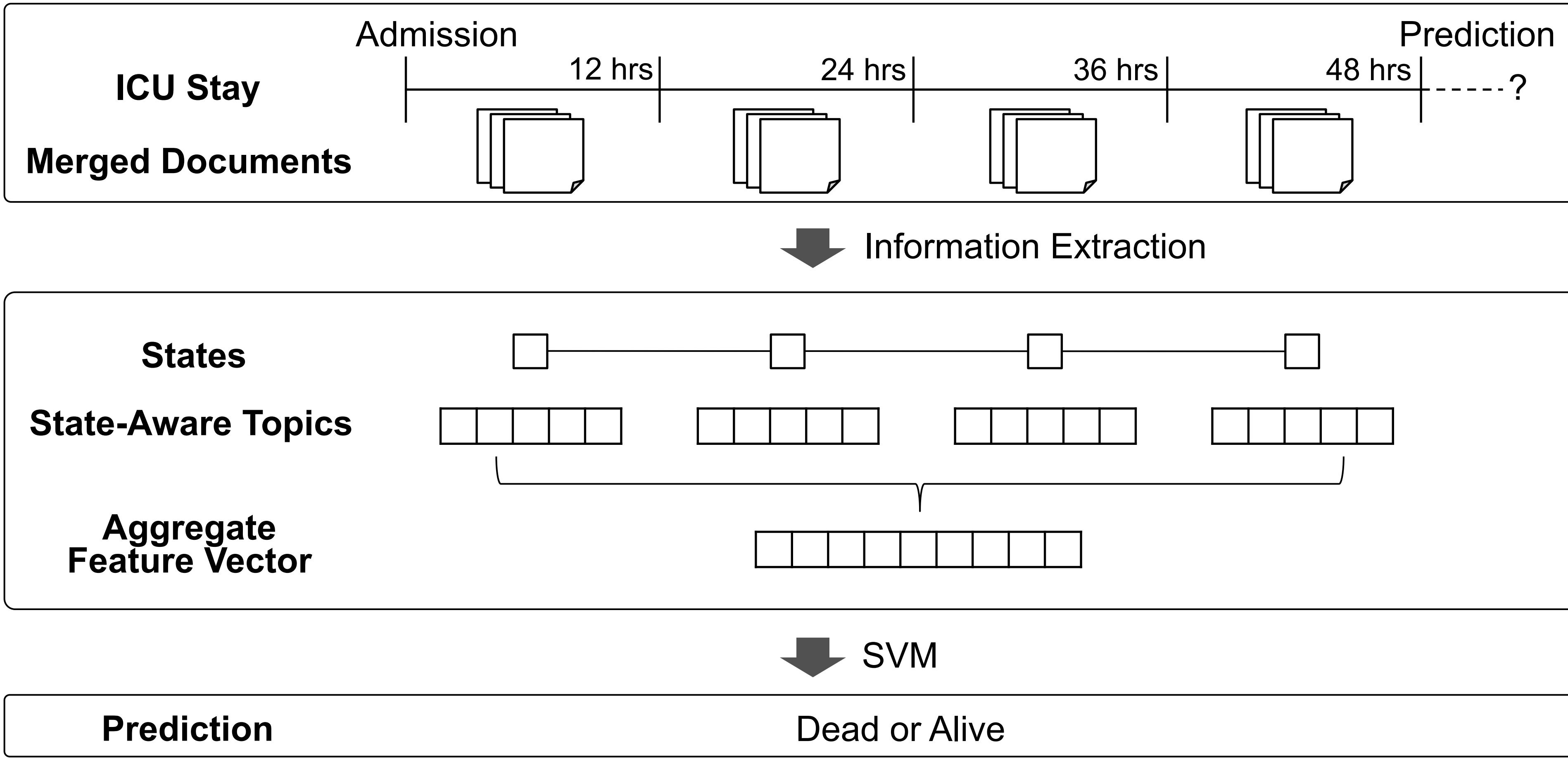
-- MIMIC2 Dataset by Physionet

Previous Study (Ghassemi et al., 2014)

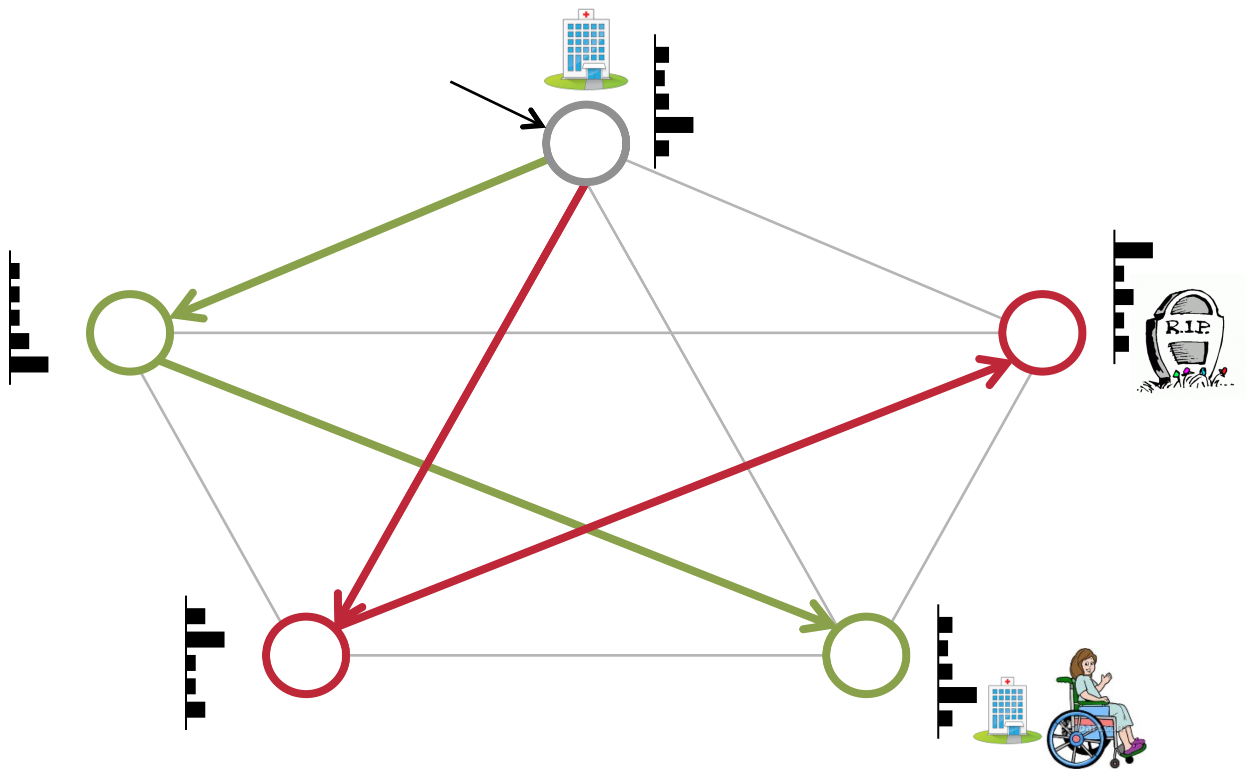


No temporal aspect is considered

This Study

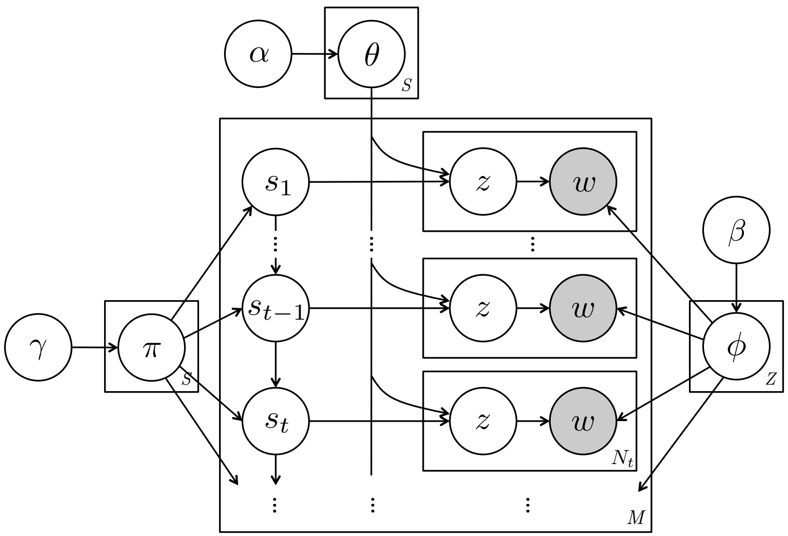

Admission
Prediction
ICU Stay
12 hrs
24 hrs
36 hrs
48 hrs
?
Merged Documents
Information Extraction
States
State-Aware Topics
Aggregate Feature Vector
SVM
Prediction
Dead or Alive

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

- ϕ_j : word distribution of each topic
- θ_c : topic distribution of each state
- $\theta^m_{t,:}$: topic distribution of each document
- π_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

one time point is 12 hrs

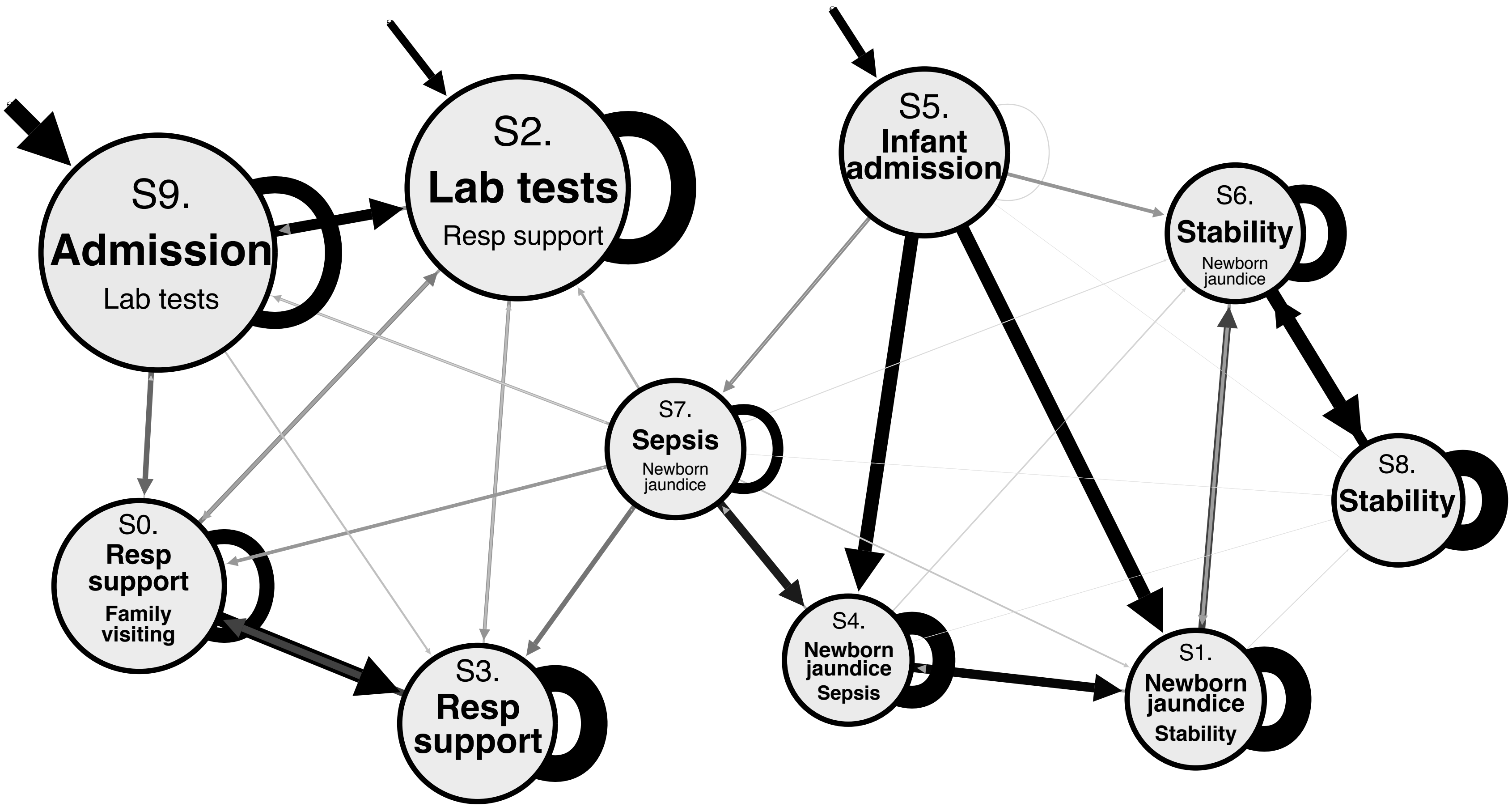
# of sequences (=# of patients)	8,808
Avg length of sequences	11 (Stddev=23)
Avg length of each document	1,548 chars (Stddev=904)

Died within	*1 day*	*1 week*	*1 month*	*6 months*	*1 year*
%	3	8	14	19	22

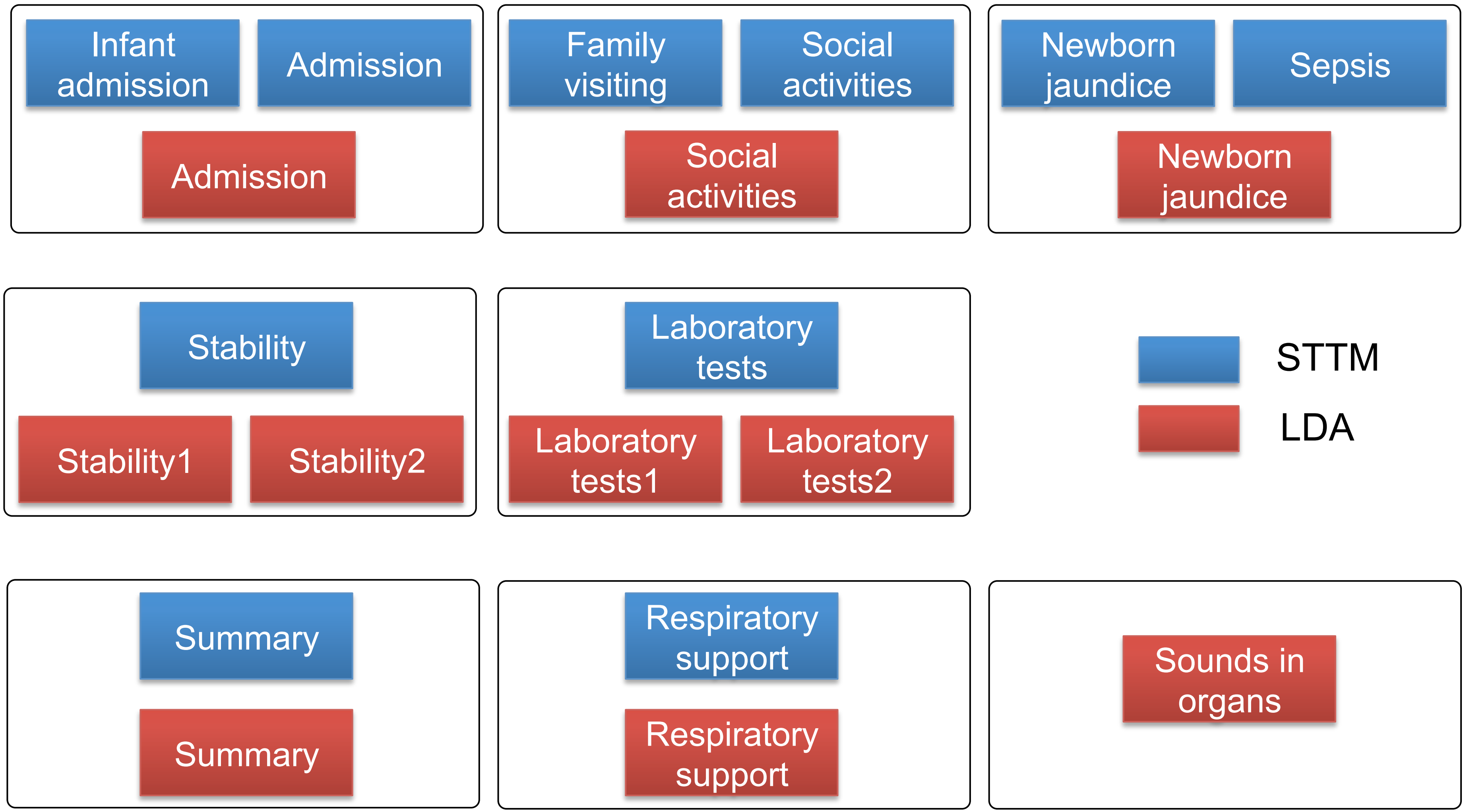
State-Aware Topics

T0. Infant admission	**T5. Social activities**
infant, nicu, delivery, cbc, normal, sepsis, maternal, born, admission, risk, baby, gbs, PHI_DATE, apgars, pregnancy	he, she, that, her, his, have, been, but, had, very, also, will, PHI_LASTNAME, if, need, would, after, about, did, they, it
Admit/transfer note Infant admitted from l&D for sepsis eval. [**Name8 (MD) 63**] MD/NNP note for hx and physical.	NPN Pt left AMA, he felt that he needed to get home to help take care of his daughters
T1. Stability	**T6. Summary**
feeds, cares, infant, well, voiding, continue, parents, feeding, active, stable, support, cc, ra, far, crib, swaddled, mom, fen	am, social, place, blood, progress, due, care, up, time, pm, fluid, npn, per, noted, cxr, id, area, cc, pain, urine, site, currently
continue to encourage po feeds. 4: G/D temps stable in an open crib. alert and active with cares. sleeps well inbetween.	NURSING PROGRESS NOTE 11 PM - 7 AM NO COMPLAINTS HR 80'S...SBP 90-100'S..
T2. Family visiting	**T7. Respiratory support**
family, patient, but, today, pts, sats, team, placed, iv, mg, bp, mask, off, afternoon, night, started, status, able, day, continues	thick, vent, secretions, peep, coarse, sputum, abg, suctioned, yellow, goal, tube, trach, remains, gtt, eyes, amts, skin
npn Pt expired at 7:40pm. Family at bedside with pt. Emotional support given to family at this very difficult time.	Resp Care: Pt continues trached and on ventilatory support with simv 600x14/+5 peep/5 psv/fio2 .4 with acceptable abg
T3. Admission	**T8. Sepsis**
ccu, micu, PHI_DATE, cath, s/p, c/o, PHI_HOSPITAL, admitted, admission, sob, floor, arrival, hct, arrived, ns, wife	npo, settings, fio2, pn, infusing, secretions, cbg, rate, vent, amp, coarse, respiratory, gas, lytes, received, ett, white, simv, cloudy
Pt transfered to [**Hospital1 22**] for ERCP which was done on arrival to MICU.	Respiratory Care Note Pt.continues on SIMV 22/5 R 34 and FIO2 33-44%. BS are coarse. Pt. sx'd for mod.
T4. Newborn jaundice	**T9. Laboratory tests**
bili, infant, cpap, baby, isolette, caffeine, servo, nested, phototherapy, enteral, mom, feeds, cc/kg/day, cc/k/d, will	pain, sbp, gtt, neuro, clear, cough, intact, oob, mg, pulses, foley, neo, bs, effect, ct, bases, given, activity, mae, lopressor, po
#6-O: under single phototherapy, bili pending this am. sl jaundiced. #4-O; temp stable in servo isolette, active and alert	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.

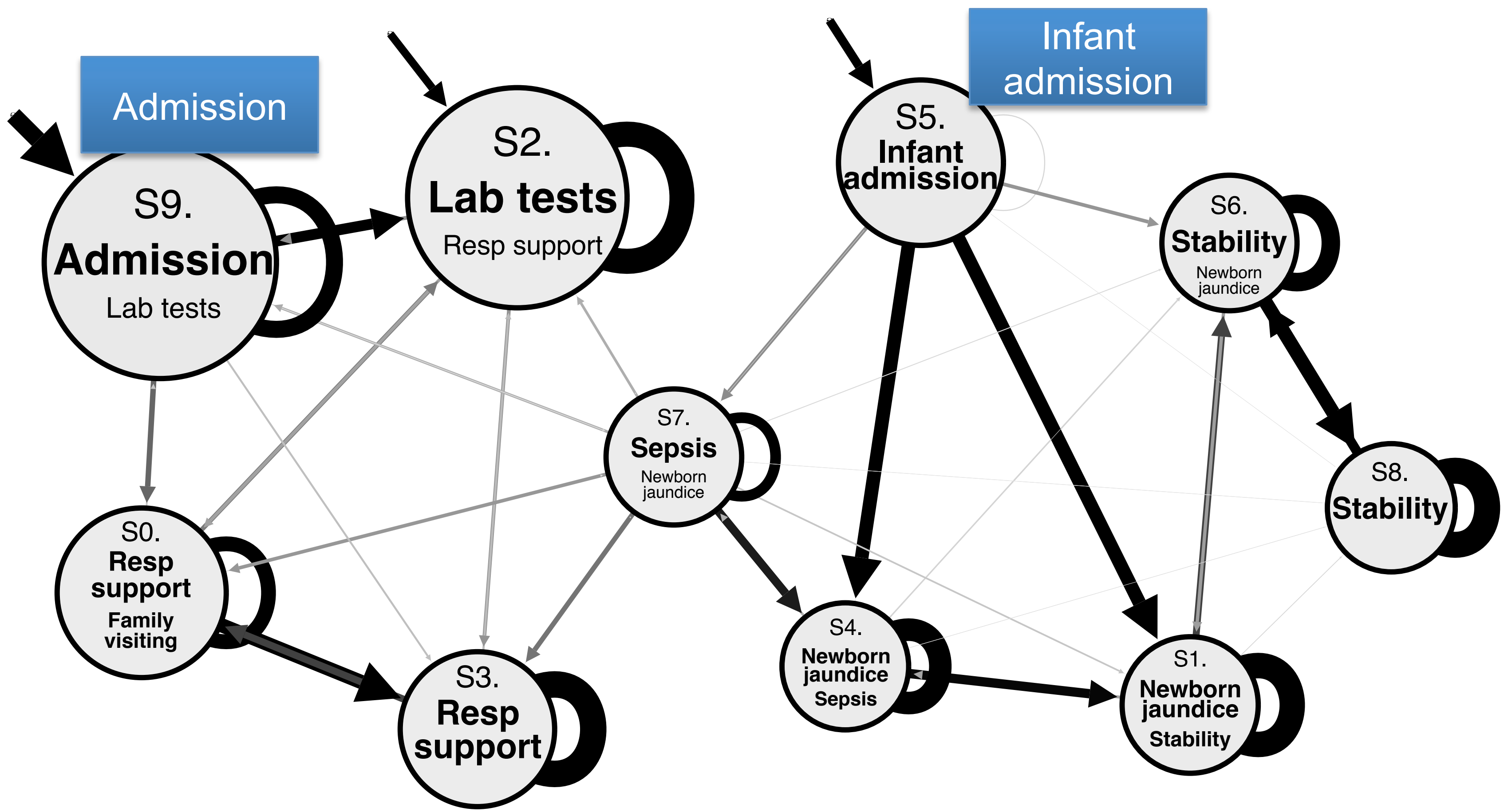
State Transitions Learned from Nursing Notes



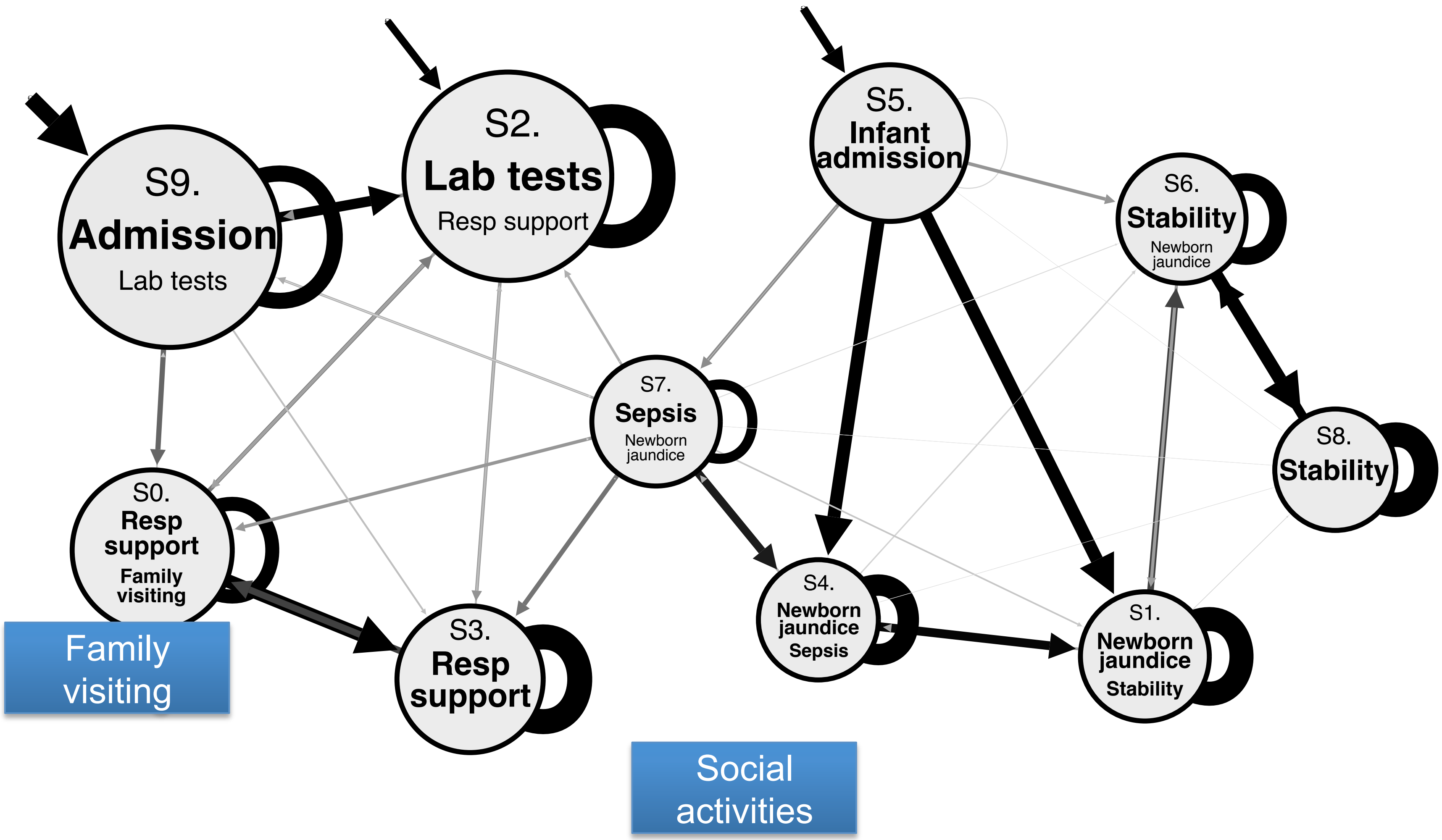
STTM vs. LDA


Infant admission
Admission
Admission
Family visiting
Social activities
Social activities
Newborn jaundice
Sepsis
Newborn jaundice
Stability
Stability1
Stability2
Laboratory tests
Laboratory tests1
Laboratory tests2
STTM
LDA
Summary
Summary
Respiratory support
Respiratory support
Sounds in organs

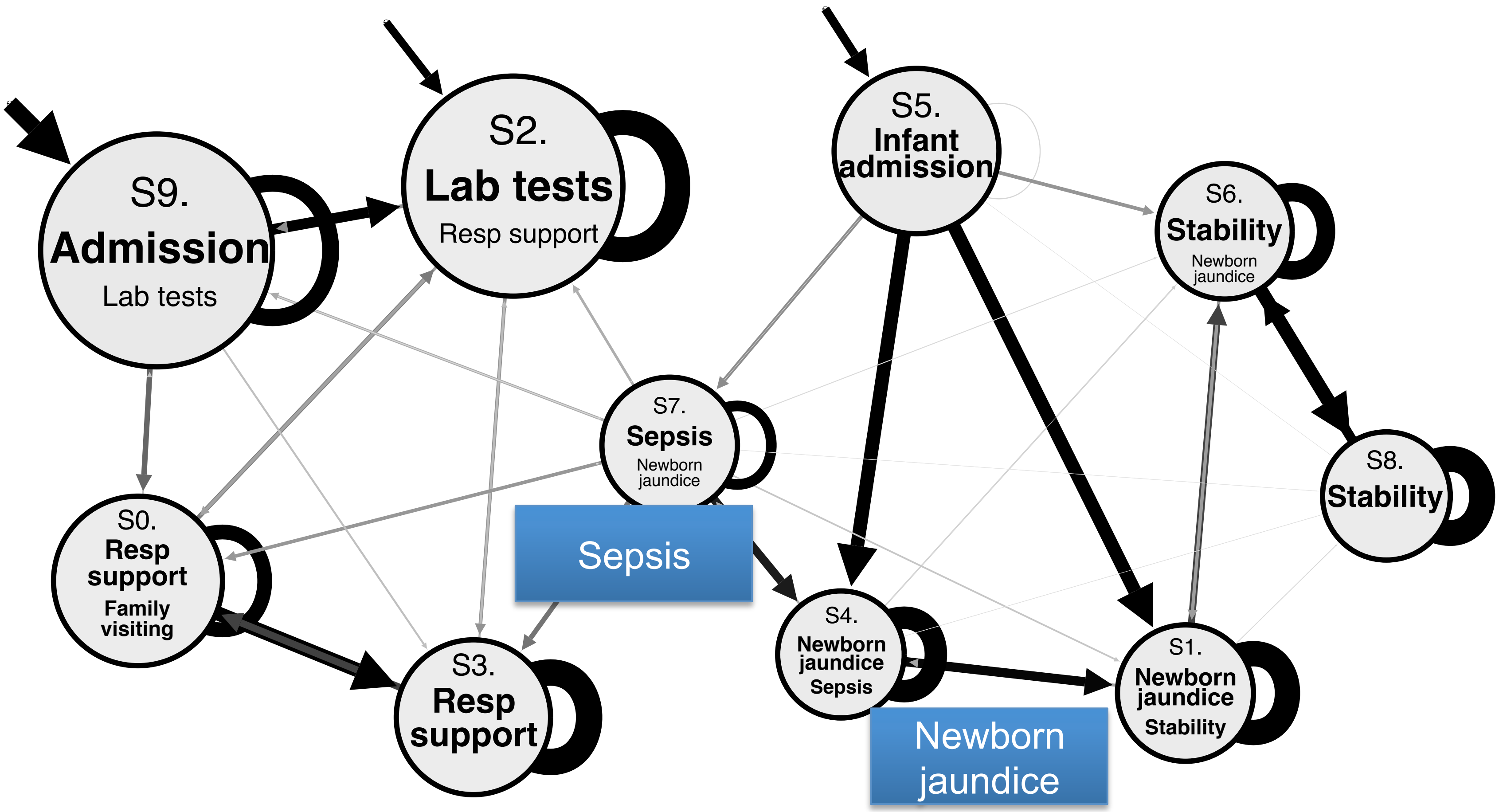
State Transitions Learned from Nursing Notes



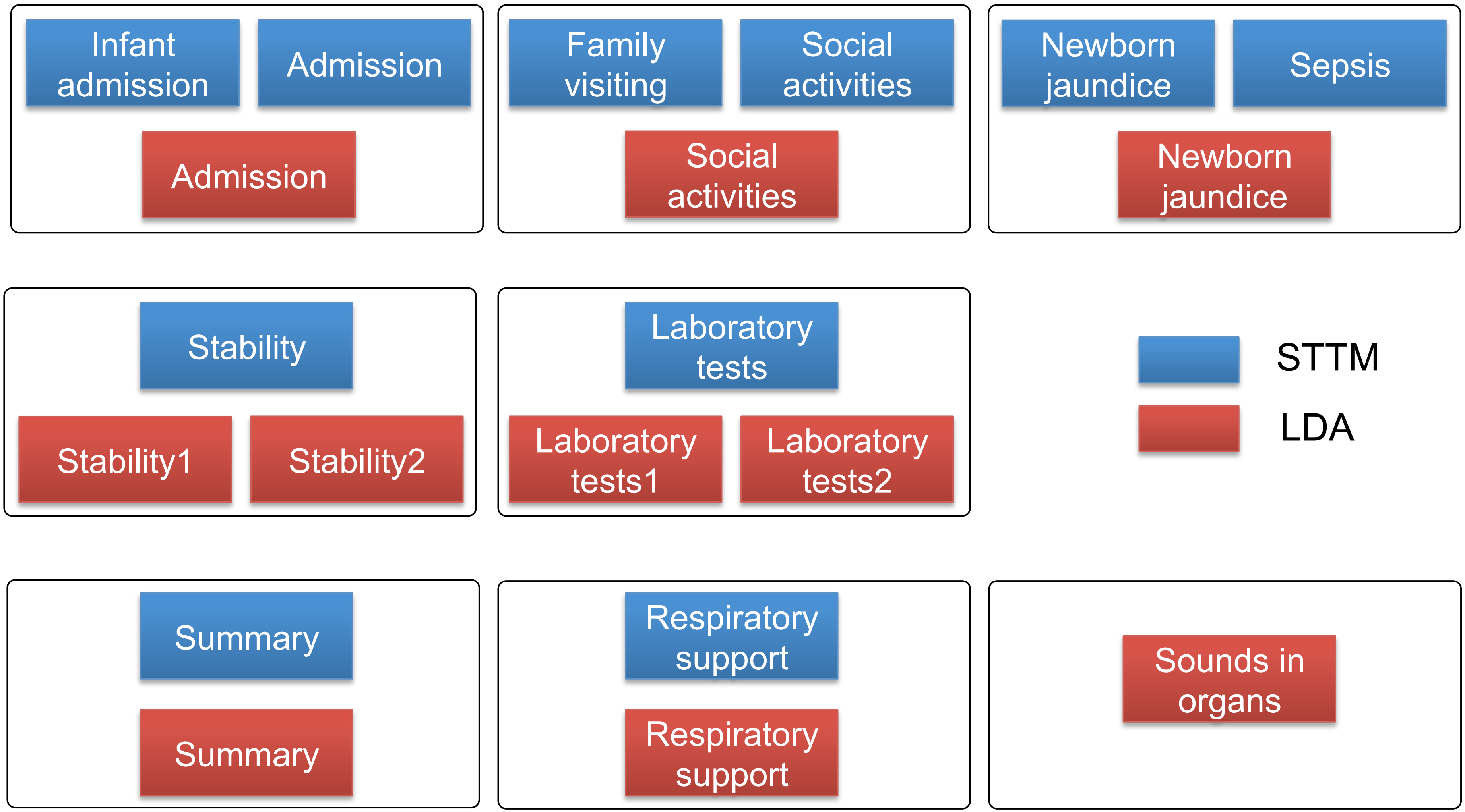
State Transitions Learned from Nursing Notes



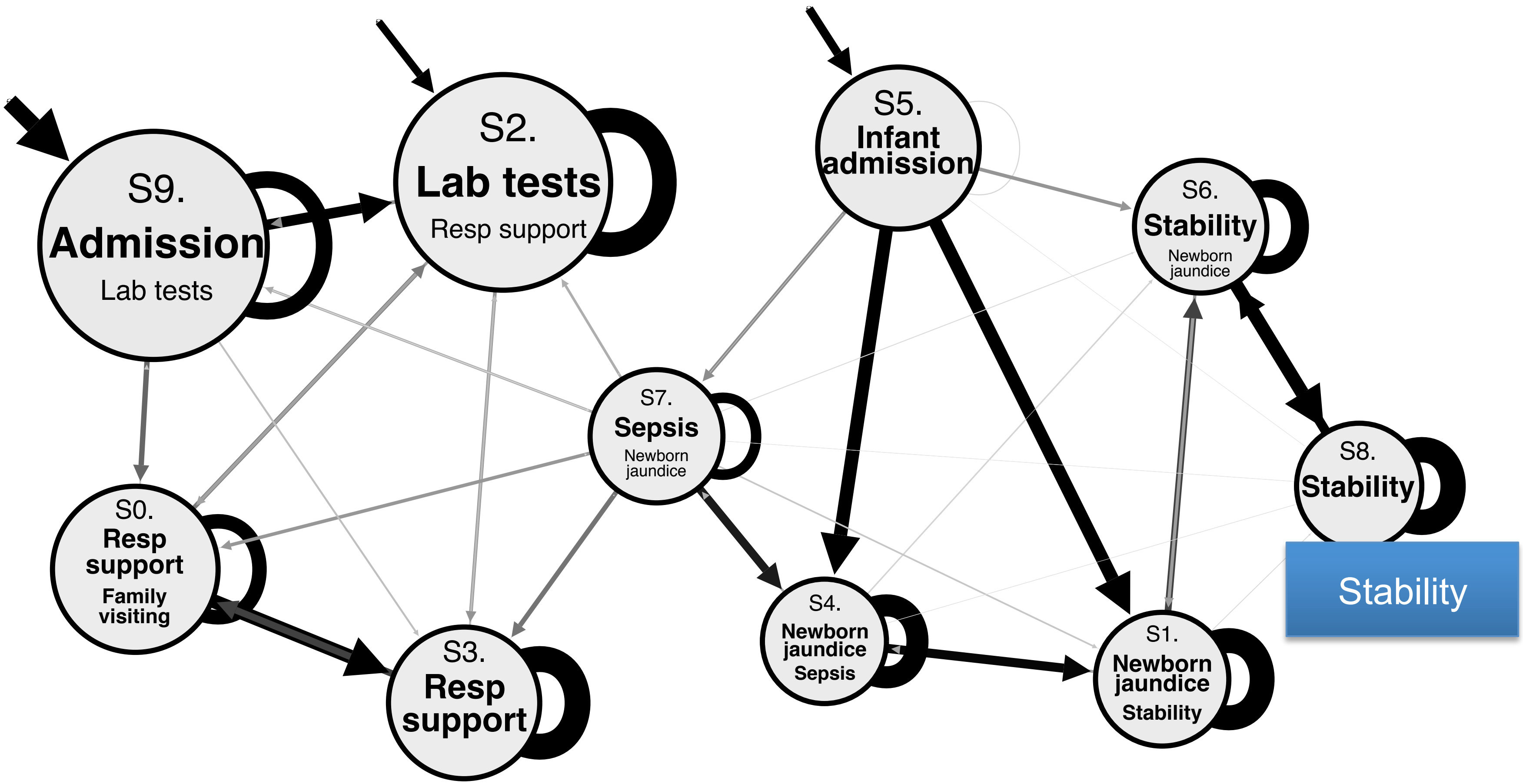
State Transitions Learned from Nursing Notes



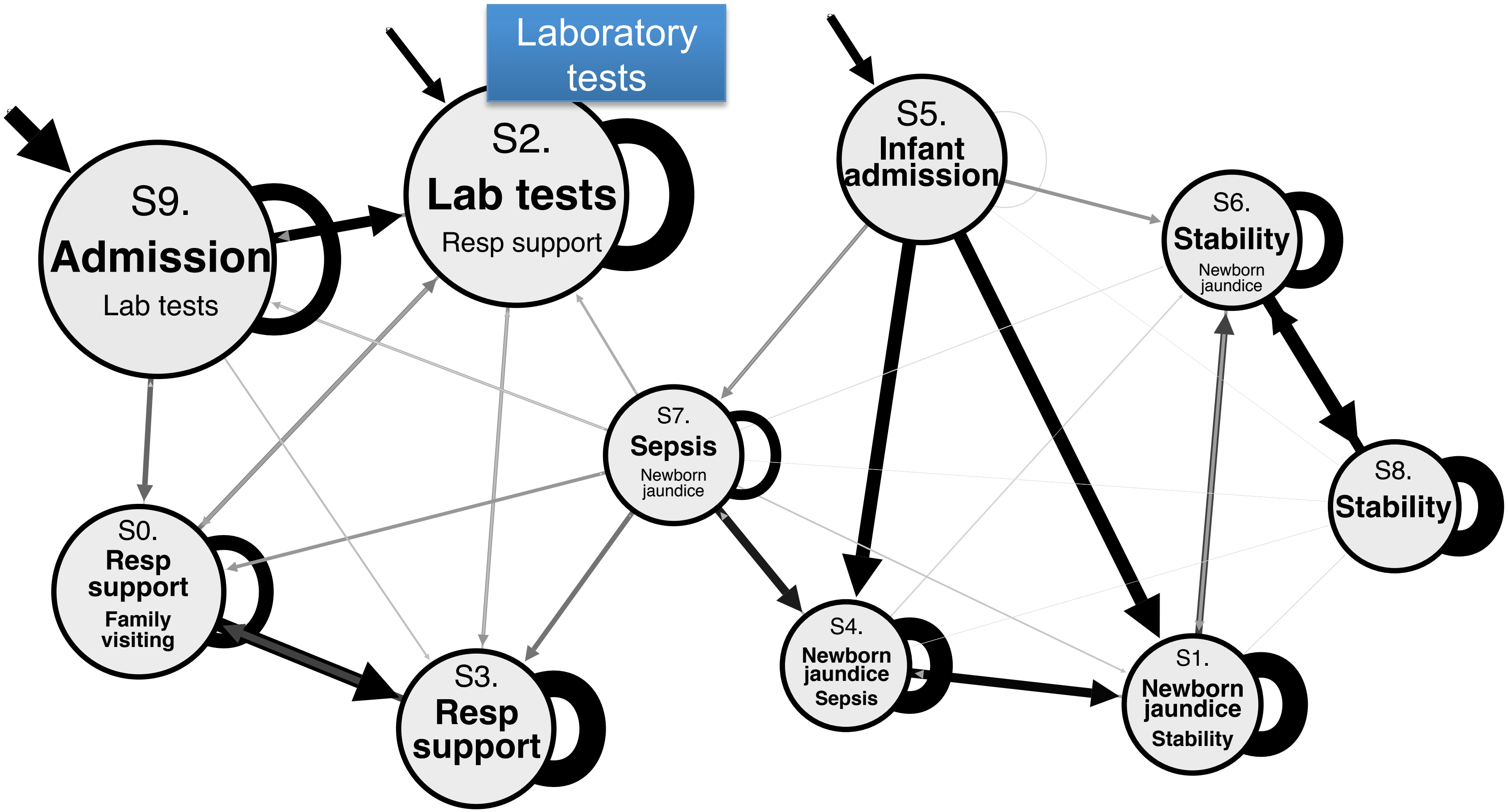
STTM vs. LDA


Infant admission
Admission
Admission
Family visiting
Social activities
Social activities
Newborn jaundice
Sepsis
Newborn jaundice
Stability
Stability1
Stability2
Laboratory tests
Laboratory tests1
Laboratory tests2
STTM
LDA
Summary
Summary
Respiratory support
Respiratory support
Sounds in organs

State Transitions Learned from Nursing Notes



State Transitions Learned from Nursing Notes

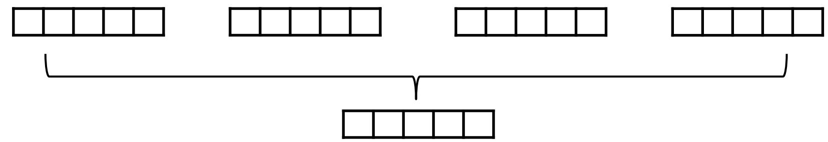


Task 2.

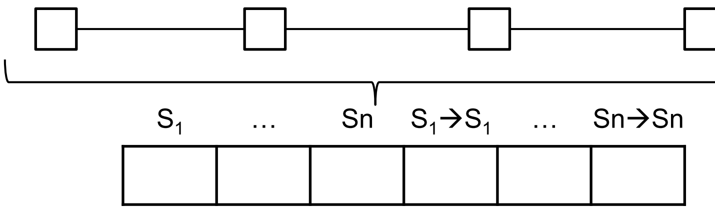
MORTALITY PREDICTION WITH TEMPORAL INFORMATION

Features

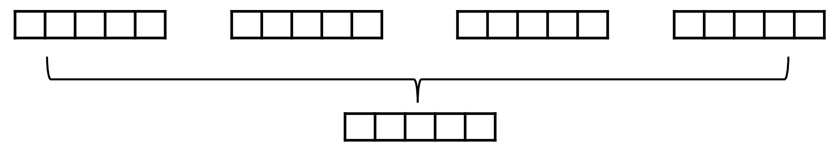
- GT
  - Standard Topics (50 topics by LDA) – Baseline



- 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

	GT	*GT+ST*	*GT+SA*	*GT+SA+*
1-Day	**0.7325**	0.7218	0.7224	0.7235
1-Week	0.7781	0.7772	**0.7811**	0.7784
1-Month	0.7820	*0.7860	*0.7866	***0.7871**
6-Month	0.7882	0.7884	***0.7921**	*0.7912
1-Year	0.7905	0.7912	0.7930	***0.7939**

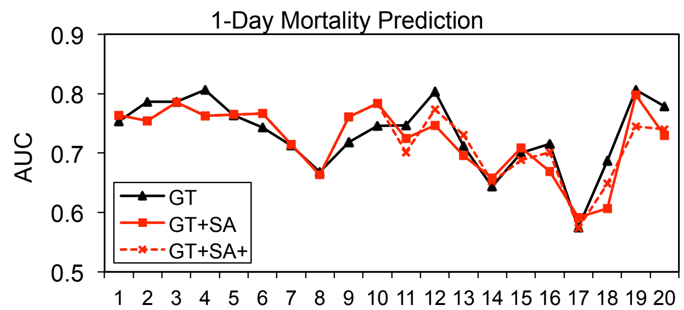
- Metric: AUC (Area Under ROC Curve)

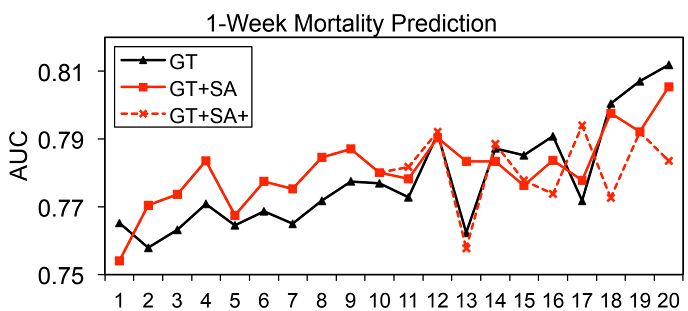
$$\mathrm{AUC} = \frac{\sum_{i=1}^{n^+} \sum_{j=1}^{n^-} \mathbf{1}\left(f(x_i) > f(x_j)\right)}{n^+ n^-}$$

- 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

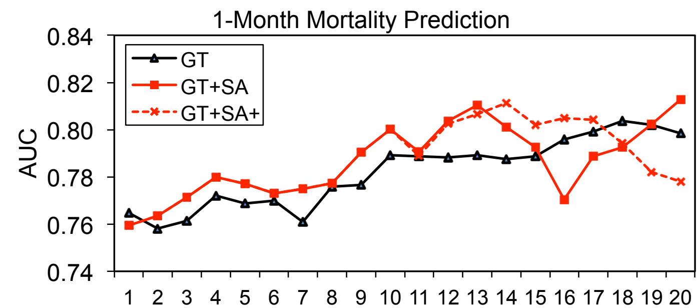
Mortality Prediction by Terms and Time Points

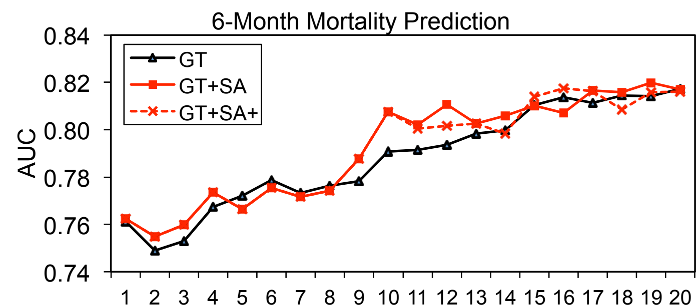
Short-Term

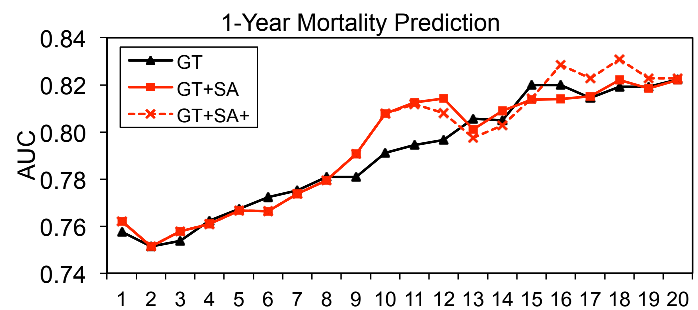




Long-Term







Task 3.

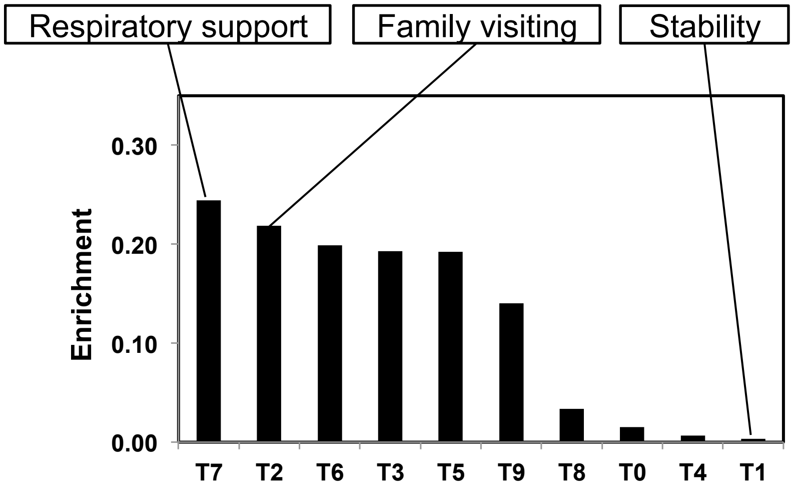
MORTALITY PREDICTION BY INDIVIDUAL FEATURES

Mortality Prediction with Individual Features

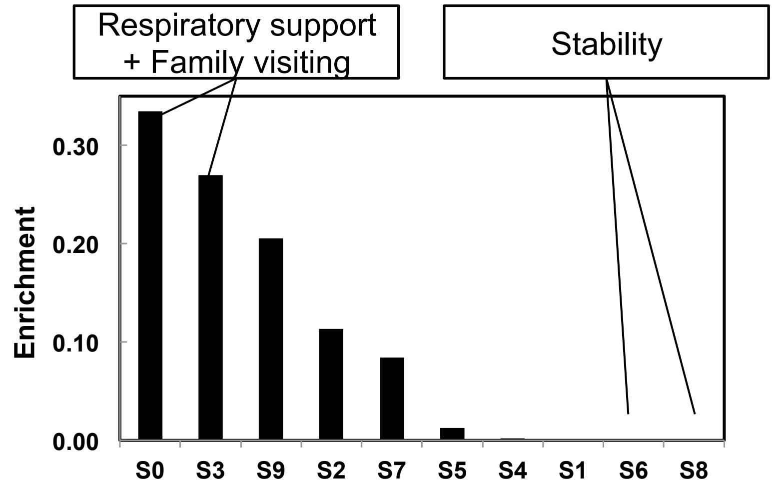
	n-grams	*10 Topics*				*50 Topics*			
		StandTopic	*StateTopic*	*StateTrans*	*StateAll*	*StandTopic*	*StateTopic*	*StateTrans*	*StateAll*
1-Day	0.563	0.709	**0.711**	0.709	0.708	***0.733**	0.705	0.703	0.691
1-Week	0.651	0.746	**0.749**	0.746	0.747	***0.778**	0.749	0.737	0.751
1-Month	0.691	0.753	0.759	0.758	***0.762**	***0.782**	0.758	0.749	0.763
6-Month	0.726	0.766	0.767	0.766	***0.771**	***0.788**	0.769	0.757	0.769
1-Year	0.732	0.772	0.771	0.769	**0.773**	***0.790**	0.776	0.760	0.775

- 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

- Chia, C.-C., & Syed, Z. (2011). Computationally Generated Cardiac Biomarkers: Heart Rate Patterns to Predict Death Following Coronary Attacks. In Proceedings of the 2011 SIAM International Conference on Data Mining (pp. 735–746). http://doi.org/10.1137/1.9781611972818.63
- Chia, C., Syed, Z., & Arbor, A. (2014). Scalable Noise Mining in Long-Term Electrocardiographic Time-Series to Predict Death Following Heart Attacks. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’14, 125–134.
- Ghassemi, M., Naumann, T., Doshi-Velez, F., Brimmer, N., Joshi, R., Rumshisky, A., & Szolovits, P. (2014). Unfolding Physiological State : Mortality Modelling in Intensive Care Units. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’14, 75–84.
- Lehman, L.-W., Adams, R., Mayaud, L., Moody, G., Malhotra, A., Mark, R., & Nemati, S. (2014). A Physiological Time Series Dynamics-Based Approach toPatient Monitoring and Outcome Prediction. IEEE Journal of Biomedical and Health Informatics, PP(99), 1. http://doi.org/10.1109/JBHI.2014.2330827
- Lehman, L., Saeed, M., Long, W., Lee, J., & Mark, R. (2012). Risk stratification of ICU patients using topic models inferred from unstructured progress notes. AMIA ... Annual Symposium Proceedings / AMIA Symposium. AMIA Symposium, 2012, 505–11. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3540429&tool=pmcentrez&rendertype=abstract
- Syed, Z., & Guttag, J. V. (2010). Identifying Patients at Risk of Major Adverse Cardiovascular Events Using Symbolic Mismatch. In Advances in Neural Information Processing Systems (pp. 2262–2270). Retrieved from http://papers.nips.cc/paper/4078-identifying-patients-at-risk-of-major-adverse-cardiovascular-events-using-symbolic-mismatch
- Syed, Z., Scirica, B. M., Mohanavelu, S., Sung, P., Michelson, E. L., Cannon, C. P., … Guttag, J. V. (2009). Relation of death within 90 days of non-ST-elevation acute coronary syndromes to variability in electrocardiographic morphology. The American Journal of Cardiology, 103(3), 307–11. http://doi.org/10.1016/j.amjcard.2008.09.099
- Syed, Z., Stultz, C. M., Scirica, B. M., & Guttag, J. V. (2011). Computationally generated cardiac biomarkers for risk stratification after acute coronary syndrome. Science Translational Medicine, 3(102), 102ra95. http://doi.org/10.1126/scitranslmed.3002557

(a) Statistics of retrieved nursing notes

# of sequences (=# of patients)	8,808
# of notes	187,808
# of merged documents	97,769
Avg length of merged documents	1,548 chars (Stddev=904)
Avg length of sequences	11 (Stddev=23)

(b) Characteristics of ICU types

	NICU	*CSRU*	*MICU*	*CCU*	*FICU*	*SICU*
# of patients	2332	2244	1918	1358	711	245
Avg. stay	10	4	5	4	5	3

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

Died within	*1 day*	*1 week*	*1 month*	*6 months*	*1 year*
%	0.03	0.08	0.14	0.19	0.22