# Topic Models for Mortality Modeling in Intensive Care Units

**Marzyeh Ghassemi** MGHASSEM@MIT.EDU

**Tristan Naumann** TJN@MIT.EDU

**Rohit Joshi** RJOSHI@MIT.EDU

**Anna Rumshisky** ARUM@CSAIL.MIT.EDU

Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139 USA

## 1. Introduction

Mortality prediction is an important problem in the intensive care unit (ICU) because it is helpful for understanding patients' evolving severity, quality of care, and comparing treatments. Most ICU mortality models primarily consider structured data and physiological waveforms (Le Gall et al., 1993). An important limitation of these structured data approaches is that they miss a lot of vital information captured in providers' free text notes and reports. In this paper, we propose an approach to mortality prediction that incorporates the information from free text notes using topic modeling.

Topic models and their variants are becoming popular in inferring the latent structure behind a collection of documents (Blei et al., 2003; Arnold, 2010). The goal of this work is to identify common topics from the unlabeled free-text observations in patients' care notes, and to determine whether the distribution of topic membership for each patient can be used as a predictive feature for mortality in-hospital, 30 days post-discharge, and 6 months post-discharge when combined with the expected mortality for each topic. Our preliminary results show that SVM classier trained on derived topics performs better than the baseline clinical features typically used for patient severity prediction in ICU.

## 2. Methods

We queried the MIMIC II 2.6 database (M. Saeed, 2011), which includes information from the EMR on 26,870 ICU patients at the Beth Israel Deaconess Medical Center (BIDMC) collected since 2001. All free text notes in a patients record prior to the first discharge were used to build a representation for each patient. This included nursing notes, physicians notes, lab results, and radiology reports but excluded discharge summaries. A binary support vector machine (SVM) classifier (Cortes & Vapnik, 1995; Chang & Lin, 2011) with a radial basis function kernel was used to predict mortality outcomes which included in-hospital, 30 days post-discharge, and 6 months post-discharge. Kernel parameters were optimized using five-fold cross validation. Patient survival data was right-censored at 1000 days post-discharge.

We compared classifier performance under the following settings:

1. A baseline representation based on the structured information typically used by physicians for coarse-grained assessment of patient severity. Namely, patient age, gender, and four samples of the Simplified Acute Physiology Score (SAPS II) (Le Gall et al., 1993): SAPS II upon admission, minimum SAPS II during the stay, maximum SAPS II during the stay, and final SAPS II.

2. A per-patient topic member ship representation using the Latent Dirichlet Allocation (LDA) model (Blei et al., 2003; Griffiths & Steyvers, 2004) with the topics derived from the top 500 words from each patient's care notes.

Mortality outcomes included in-hospital, 30 days post-discharge, and six months post-discharge. Patients were excluded if they had fewer than 100 relevant words, did not have SAPS II scores available, or were not adults. This resulted in 19,807 patients included: 13,867 (70%) were randomly assigned as a training set for topic model creation, and 5,940 (30%) were selected as a test set. We then used five-fold cross validation for evaluating baseline feature and topic feature performance.

*Table 1.* Example of Enriched Topics

| | Topic | Top Ten Words | Possible Topic |
|---|---|---|---|
| In-Hospital Mortalities | 41 | family pt neuro name status eyes care remains movement pupils | End-of-life care requested |
| | 28 | vent pt abg intubated peep remains secretions resp ac sedated | Respiratory failure with intubated patient |
| | 44 | gtt insulin map remains line heparin levo ptt fluid levophed | Severe hypotension by sepsis-induced systemic vasodilation |
| | 37 | pt noted cc hr bp cchr am micu urine bs | Many physiological parameters noted, corresponding to critical state |
| Near-term Survival | 17 | resp pt sats cough mask sob lasix neb nc wheezes | Chronic respiratory infection |
| | 2 | status iv mg times commands agitated ativan neuro bed agitation | Agitated patient with dimension or risky behaviors |
| | 9 | catheter wire name sheath french advanced procedure contrast stent placed | Ventricular stent placed |
| Long-term Survival | 26 | co pain neuro denies oriented po alert resp diet cv | Responsive patient |
| | 32 | gtt mg lasix sbp am bs neuro lopressor bp po | Non-lethal low blood pressure (orthostatic, dehydration, etc.) |
| | 21 | fracture trauma fractures left injury sp clip fx reason ap | Non-lethal trauma |
| | 13 | chest cabg sp coronary post pain artery pa pleural wires | Cardiovascular surgery |
| | 45 | pt neo pain gtt neuro plan bs wean resp sbp | Management routine e.g. diabetes, physiotherapy, etc. |

## 3. Results

Topic feature area under the ROC curve (AUC) performance was consistently higher than the baseline performance using the age, gender, and the four SAPS II scores for all three outcomes (mortality in-hospital, 30 days post-discharge, and 6 months post-discharge). Table 2 shows topic feature performance with 50 topics compared against the baseline. The data contains a limited number of mortalities (13.7%, 15.4%, and 21.8% respectively), resulting in a classification bias which induces low sensitivity, especially for the baseline model. While the specificity of both the models are similar, the topic model achieves marginally better AUC but much higher sensitivity, suggesting that topics learned are enriched for mortality.

To further investigate the validity of derived topics for outcome predictions, we performed qualitative analysis of the topic distribution for the following outcomes: (1) in-hospital mortality, (2) near-term mortality (death within 1000 days post-discharge), and (3) long-term survival (survival 1000 days post-discharge). To examine which topics, if any, are more common in patients with specific outcomes, the median proportional topic membership for each subgroup of patients was calculated.

Figure 1 shows the median topic membership distributions. Each patient's proportional membership in a given topic is based on the percentage of words which are drawn from that topic. We follow the standard topic modeling scheme to describe each patient's record as a probability mixture of topics that sums to one. For each of the three groups (in-hospital, near-term mortality, and long-term survival), we then calculate a median proportional membership by taking the median topic membership over all the patients in that group. For example, if the "in-hospital" group contains $k$ patients, then we take the median topic membership probabilities for each topic over these $k$ patients. The 30 days post-discharge and 6 months post-discharge outcome groups had similar topic proportions, and so were combined in Figure 1 for the purposes of illustration.

Table 1 shows some of the top topics for each outcome; note that the topic numbers referenced in Table 1 correspond to those shown in Figure 1. Qualitative analysis of these topics suggested that they align well with the major known causes of ICU admission and mortality, as noted in the rightmost column. Figure 1 shows the proportion of outcomes associated with each topic. Some topics were clearly associated with in-hospital or near-term mortalities, while others only

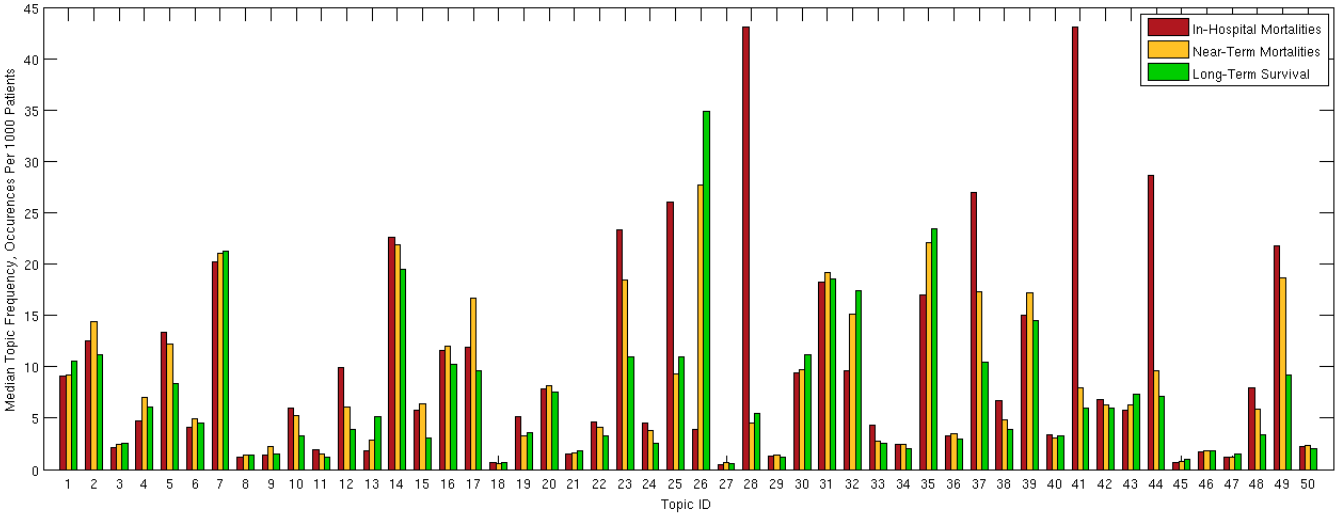


*Figure 1.* Median Topic Membership Distributions per 1000 Patients. The total height of a colored bar corresponds to the median number of words that are accounted for by that topic in the group which corresponds to the color. E.g. in 1000 patients, the median record in the "in-hospital mortality" category will contain 43 words generated by topic 28, compared to 5 words in the median record of those alive past 1000 days. Note that numbers are estimated on median percentages for each group.

occurred in records of long-term survivors.

*Table 2.* Model Classification Performance

| Outcome | Features | AUC | Sens. | Spec |
|---|---|---|---|---|
| In-Hospital | Baseline | 0.839 | 0.069 | 0.998 |
| Mortality | Topics | 0.855 | 0.465 | 0.985 |
| 30 Day | Baseline | 0.730 | 0.111 | 0.992 |
| Mortality | Topics | 0.754 | 0.484 | 0.983 |
| 6 Month | Baseline | 0.711 | 0.135 | 0.980 |
| Mortality | Topics | 0.781 | 0.438 | 0.966 |

## 4. Conclusion and Future Work

New advances in severity scoring in ICU are required to provide improved therapeutic effectiveness and cost-effective ICU care. We showed that topic models have a potential to create natural high-level categories of conditions which align with known mortality outcomes and known causes of ICU death. We also found that topics have interesting structure with respect to the mortality time of patients. Preliminary results suggest that the Latent Dirichlet Allocation model with bag of words does outperform SAPS-II baseline for the 6-month mortality category. However, minimal performance gain for other categories sugggests a need for more advanced NLP processing or a different variant of topic modeling for severity prediction from clinical text. Our results show that the real-time text can act as a substitute for other ICU scores, such as SAPS II, which tend to become less reliable as patients spend longer amounts of time in the ICU.

Our model used a text representation that did not incorporate information about linguistic structure. The model should benefit from adding negation and modality detection, named entity recognition, text sectioning, etc. Further, its likely that our model will benefit from using better priors for topic modeling. In our ongoing work, we are exploring the generalizability of topic models such that they can be used as a standard way for mortality comparison among different hospitals. In future, we plan to extend our proposed technique to do more advanced patient severity modeling, automated disease surveillance and per-patient summaries for quick evaluations.

## Acknowledgments

This research was made possible by funding from grant 2R01 EB001659 from the National Institute of Biomedical Imaging and Bioengineering (NIBIB), NIH; grant T15LM007092 from the National Library of Medicine, NIH; Xerox Fellowship Program initiated in February, 2007, Xerox Foundation.

## References

Arnold, C.W. et al. Clinical case-based retrieval using latent topic analysis. In *AMIA Annual Symposium Proceedings*, volume 2010, pp. 26. AMIA, 2010.

Blei, D., Ng, A., and Jordan, M. Latent dirichlet allocation. *JMLR*, 3(5):993–1022, 2003.

Chang, Chih-Chung and Lin, Chih-Jen. LIBSVM: A library for support vector machines. *ACM TIST*, 2: 27:1–27:27, 2011.

Cortes, Corinna and Vapnik, Vladimir. Support-vector networks. *Machine Learning*, 20:273–297, 1995. ISSN 0885-6125.

Griffiths, T. and Steyvers, M. Finding scientific topics. In *PNAS*, volume 101, pp. 5228–5235, 2004.

Le Gall, J.R., Lemeshow, S., and Saulnier, F. A new simplified acute physiology score (saps ii) based on a european/north american multicenter study. *JAMA*, 270(24):2957–2963, 1993.

M. Saeed, et al. Multiparameter Intelligent Monitoring in Intensive Care II: A public-access intensive care unit database. *Critical Care Medicine*, 39(5):952–960, May 2011.