Interpretable Topic Features for Post-ICU Mortality Prediction

Yen-Fu Luo, MS, Anna Rumshisky, PhD
University of Massachusetts Lowell, Lowell, MA

Abstract

Electronic health records provide valuable resources for understanding the correlation between various diseases and mortality. The analysis of post-discharge mortality is critical for healthcare professionals to follow up potential causes of death after a patient is discharged from the hospital and give prompt treatment. Moreover, it may reduce the cost derived from readmissions and improve the quality of healthcare.

Our work focused on post-discharge ICU mortality prediction. In addition to features derived from physiological measurements, we incorporated ICD-9-CM hierarchy into Bayesian topic model learning and extracted topic features from medical notes. We achieved highest AUCs of 0.835 and 0.829 for 30-day and 6-month post-discharge mortality prediction using baseline and topic proportions derived from Labeled-LDA. Moreover, our work emphasized the interpretability of topic features derived from topic model which may facilitates the understanding and investigation of the complexity between mortality and diseases.

1. Introduction

Post-discharge management is one of the important aspects in current healthcare system. For high-risk patients, and especially for the intensive care unit (ICU) patients, it is critical to understand and prevent possible complications and problems which may lead to a patient's death after being discharged from the hospital. The present work focused on mortality prediction of high-risk ICU patients. In our patient cohort, the post-discharge mortality for 30-day and 6-month are 3.4% and 9.5% respectively. There has been a lot of recent interest in mortality prediction in general and post-ICU mortality prediction in particular[1-3]. However, many of the state-of-the-art methods use "black box" predictive models which can not provide any explanation for practitioners as to why a particular patient may be at risk after discharge. In this paper, our goal is two-fold: develop novel methods that can both accurately predict mortality and at the same time create a *transparent predictive model* that can be easily understood and therefore actionable by the providers.

SAPS-II[4], APACHE-II[5], and SOFA[6] scores are commonly used in ICU mortality prediction[1-2, 7-10]. In addition to structured data and derived severity scores, we build a mortality prediction model that incorporates features derived from unstructured medical notes. We use Multiparameter Intelligent Monitoring in Intensive Care (MIMIC II)[11] database. The narrative provider notes from MIMIC II give detailed descriptions of symptoms, diagnosis, surgery, medicine, and treatments. It is highly informative but in the form of free-text. Our goal is to capture clinically relevant information and patterns identified and summarized by healthcare providers in order to leverage them in *transparent prediction*.

There have been recent attempts to use Bayesian topic modeling techniques to improve mortality prediction using narrative notes[1-3, 7]. In topic modeling, each document is represented as a probability distribution over a set of topics and each topic is modeled as a probability distribution over a set of words. Although topic-based features have been used in literature to improve outcome prediction, the topics themselves are flat word collections that need to be examined by domain expert in order to assign a clinical interpretation.

Although the derived topics show some degree of interpretability[12], human annotators are prone to assigning meaning to topics or word clusters even in cases when such word collections are not coherent. In this work, we propose a method to automatically define interpretable topics. To make a topic itself interpretable and clearly definable based on domain knowledge, we used the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes as topics namely labels in Labeled-LDA to guide topic model learning and extract understandable topic feature representations from medical notes. In addition, we examined the feasibility of using topic features derived from Labeled-LDA for post-discharge mortality prediction.

2. Related Work

One of the common approaches to using narrative notes for clinical outcome prediction is to extract clinically-relevant concept and relations using information extraction techniques and use them as features in predictive models. A number of medical concept extraction systems, including rule-based systems (MetaMap[13], MedLEE[14]

cTAKES[15], etc.) and machine learning-based systems (CliNER[16], RapTAT[17]) have been used for this task[18-23]. However, such systems require a substantial amount of human labor in order to produce accurate results, either for rule construction and keyword selection, or for text annotation and feature engineering required for the supervised machine learning. As a result, shifting between different types of clinical notes, or between different institutions, requires a substantial overhead in order to achieve domain adaption. This is additionally compounded by continuous changes in medical terminology, introduction of new medication brands, and so on.

There has been a number of recent attempts to bypass this problem by using unsupervised methods that rely on topic modeling to extract topic features from clinical narrative text in order to improve the prediction of in-hospital and post-discharge mortality for ICU patients. Ghassemi et al.[1] reported AUCs of 0.754 and 0.781 for 30-day and 6-month post-discharge mortality prediction. Ghassemi et al.[2] also reported an AUC of 0.818 for 30-day post-discharge mortality prediction using retrospective topic + derived features model. Lehman et al.[7] combined medical concepts extracted from medical notes with topic model for ICU in-hospital mortality prediction. Jo et al.[3] used state transition topic model for incorporating temporal information and reached an AUC of 0.792 for 6-month post-discharge mortality prediction.

Ideally, predictive models for mortality should be customized for different patient groups, based on primary diagnosis and other patient characteristics. Nori et al.[24] combined the hierarchy of ICD-10 into mortality prediction and divided the general prediction model into multi-task/multi-disease learning problem. Makar et al.[25] also incorporated ICD-9-CM codes for short-term mortality prediction of elderly patients.

Topic models such as the Latent Dirichlet Allocation (LDA)[26] and Hierarchical Dirichlet Process (HDP)[27] are widely used to explore coherent topics within large text corpora. HDP is a nonparametric Bayesian approach which does not require specifying the desired number of topics. Arnold et al.[12] showed the interpretability of topic model from a physician's perspective. Although their conclusions support using topic features in a prediction task, identifying high quality topics may also require a labor-intensive topic evaluation by domain experts, in order to determine the optimal parameter settings (i.e., the number of topics in LDA or the concentration hyperparameter in HDP). Incorporating domain knowledge into the topic learning in the way implemented in the present work helps to address both the problem of customizing predictive models for different patient categories and to improve topic interpretability.

3. Methods

In the present work, we propose to incorporate domain knowledge into topic learning using Labeled-LDA[28] with ICD-9-CM codes as labels. Labels are equivalent to topics or ICD-9-CM codes in our setting, and each document may be assigned multiple labels. In LDA, all documents contribute to all topics in the learning of the topic model. In Labeled-LDA, a subset of documents with the corresponding label is used to infer word distributions for a topic. The benefit of using ICD-9-CM codes as labels in Labeled-LDA is two-fold. First, the clinical notes from a given patient's record contribute only to a subset of topics corresponding to the ICD-9-CM code assignments for that patient. Second, topic interpretability is achieved through a combination of the ICD-9-CM code definition and the top words for a given topic. At the training stage, we incorporate ICD-9-CM codes to guide the Labeled-LDA model learning. However, since ICD-9-CM codes are not available at the time of discharge, they can not be included as features in the predictive models directly. Using Labeled-LDA approach allows us to bypass this issue, since at the prediction stage, inferring topic proportions corresponding to different ICD-9-CM codes does not require one to have the ICD-9-CM codes available.

3.1 Patient Selection

We used MIMIC II database. The database contains physiological signals, vital signs, medical notes, and other structured data from several ICUs, including medical, surgical, coronary care, and neonatal. This data was collected between 2001 and 2008 at Boston's Beth Israel Deaconess Medical Center (BIDMC). The database contains over 25,000 patients including around 20,000 adults and 5,000 neonates.

Since the factors related to mortality differ substantially for neonates, only patients in the adult age group were selected. The patients without discharge summary were excluded from the cohort, since discharge summary is essential for building the prediction model for post-discharge mortality. We also excluded patients without the first SAPS II score. All available clinical notes, including nursing, physician, radiology, and the discharge summary of the patient's first hospital stay were collected. We identified and removed 11% duplicate medical notes. The resulting cohort consisted of 18,412 patients with 400,494 notes. The patient data was randomly split into 80% for training set and 20% for testing set.

3.2 Preprocessing and Tokenization

Each note was processed by using the SPECIALIST Lexicon LRABR table to preserve medical abbreviations and acronyms; this was followed by whitespace-based tokenization, and the removal of stopwords from the Onix stopword list. Term frequency was generated by aggregating word frequency for each patient. The top 500 most informative words were selected for each patient, based on the TF-IDF score[29] for each patient included in the training data. This resulted in the overall vocabulary containing 151,772 words.

3.3 LDA and Labeled-LDA

Knowledge Based Topic Models (KBTM) were developed to guide topic model learning by incorporating domain knowledge. Andrzejewski et al.[30-31] demonstrated the use of Dirichlet forest priors and first-order logic in order to create must-links and cannot-links between words which encode domain knowledge during model learning. This solution requires domain experts to encode knowledge used to create constraints. We propose using Labeled-LDA as an alternative. Labeled-LDA was designed to analyze the text of the web pages that may be annotated by users in a community portal. Each page may have multiple labels associated with the topics of the page (such as arts, politics, physics, religion, alaska, etc.), assigned by the readers. In Labeled-LDA, word distribution of a topic is inferred based on a subset of the corpus with the corresponding label. Therefore, the inferred top keywords of a topic are associated with the subject of the label.

We adapted this model to the task of *transparent outcome prediction*, using the ICD-9 diagnostic codes assigned to each patient as labels, with the two-fold goal of guiding the topic learning and improving interpretability of the resulting topics. ICD-9-CM is based on the World Health Organization's Ninth Revision, International Classification of Diseases (ICD-9). Until October 2015, diagnoses and procedures associated with hospital utilization in the United States were recorded using official ICD-9-CM codes. Based on multiple procedures and treatments during a patient's hospital stay, multiple ICD-9-CM codes are assigned by trained healthcare professionals.

Although there is some disagreement on the viability of using ICD-9 codes in predictive tasks due to the diagnostic codes being assigned exclusively for billing purposes, they do provide an expert-generated authoritative source of annotation for each record, which can be reasonably assumed to represent high-level domain knowledge. The topics obtained by applying Labeled-LDA to medical notes of ICU patients with ICD-9-CM codes as labels may be interpreted as providing a description of sorts to the corresponding code, which can be easily verified against the ICD-9 code definition. Under this interpretation, the topic proportions also represent the extent to which a particular diagnosis or procedure is associated with the given hospital stay.

Rather than using the raw ICD-9-CM codes, we used the ICD-9-CM hierarchy with 180 upper-level codes in order to reduce the sparsity of ICD-9-CM code assignments. Comparing to the LDA model that uses all documents to infer topic proportions and word distributions, the Labeled-LDA model only uses a subset of documents for each topic inference. Since training a topic model requires sufficient data to produce coherent topics, we only considered labels with minimum frequency 50, 100, 200, and 400. The resulting number of labels are 111, 94, 79, and 59 respectively. Following Ghassemi et al[1-2], we used 50 topics for training a regular LDA model. We sampled topic proportions for each patient in the training data after using 2,500 iterations during model learning, and the resulting model was used to obtain topic proportions for each patient in the test data. We used default hyperparameter settings in both LDA and Labeled-LDA models.

3.4 Mortality Prediction

We retrieved age, gender, SAPS-II scores, Elixhauser Comorbidity Index[32], the text of the medical notes, and the ICD-9-CM code assignments for each patient's first hospital stay recorded in the MIMIC II database. The topic model was used to infer topic proportions for all medical notes in a patient's record at the time of discharge. This information, together with the obtained topic proportions, were used as real-valued features in a predictive model. A Support Vector Machine (SVM) model[33] with radial basis function (RBF) kernel was trained and used to predict 30-day and 6-month post-discharge mortality.

For 6-month post-discharge mortality prediction, we used three feature settings: (1) baseline features included age, gender, SAPS-II score at admission, minimum SAPS II score, maximum SAPS II score, and 30 Elixhauser Comorbidities, (2) baseline features and 50 topic proportions derived from regular LDA, and (3) baseline features and topic proportions derived from Labeled-LDA. For the 30-day prediction model, we excluded minimum SAPS-II score, maximum SAPS-II score, and ICD-9-CM derived 30 Elixhauser Comorbidities from the baseline features.

The reason is that the assignment of ICD-9-CM codes is usually finalized within 2 weeks after a patient is discharged from the hospital, and therefore is not available at the time the prediction needs to be made.

In our patient cohort, 3.4% and 9.5% of the patients died within 30-day and 6-month post-discharge respectively. Because of the highly imbalanced data, we subsampled negative class to generate dataset with 20% positive and 80% negative class for training. In addition, we penalized misclassification of positive class by assigning higher class weight in SVM. The optimal cost and gamma parameters were determined in 5-fold cross-validation over the training data against ROC-AUC.

4. Results

4.1 Topic Interpretability

Baseline + Labeled-LDA with 111 labels achieved highest AUC in both 30-day and 6-month mortality prediction. To illustrate the topics derived with this Labeled-LDA model, we show 10 most and 10 least frequent ICD-9-CM codes in Table 1, along with their definitions, and top 20 words. The results suggest the consistency between ICD-9-CM code's definition and the corresponding keywords. For example, the top words for the "hypertensive disease" topic include 'chest', 'cabg', 'artery', 'coronary', etc. Another example topic, labeled "complications occurring mainly in the course of labor and delivery" is associated with the words 'uterine', 'bleeding', 'vaginal', 'delivery', 'abd', and 'hct'.

Table 1. List of top 20 words learned from Labeled-LDA and its corresponding ICD-9-CM definition. Top 10 and bottom 10 entries are most and least frequent ICD-9-CM code in our dataset. Frequency of 111 ICD-9-CM codes are listed in Table 3.

ICD-9-CM	Definition (above) / Keywords (below)
401-405	Hypertensive disease
	tablet chest left mg po sig pt daily reason cabg artery sp refills disp namepattern clip pain date day coronary
420-429	Other forms of heart disease
	pt mg patient hr chest resp left lasix gi po stable pain gu neuro gtt bp bs day cv plan
270-279	Other metabolic and immunity disorders
	patient mg pt chest day left artery pain po stable coronary cabg status discharge history date post namepattern clip examination
410-414	Ischemic heart disease
	mg pt patient cath tablet pain left cardiac chest po hospital hr artery ccu discharge coronary history normal namepattern daily
249-259	Diseases of other endocrine glands
	pt patient mg insulin day blood hr po pain discharge units bs diabetes history namepattern hospital gtt admission pm doctor
510-519	Other diseases of respiratory system
	pt hr resp vent remains cc secretions thick care tube plan bs neuro trach cont mg gi noted yellow abg
996-999	Complications of surgical and medical care, not elsewhere classified
	pt tube resp left hr chest plan neuro remains cc vent bs cont reason noted clip abd gi sp care
280-285	Anemia
	pt tablet mg blood po hct sig daily discharge pm doctor namepattern md patient pain day history gi admission hospital
780-789	Symptoms
	patient pt contrast ct head left clip seizure reason normal pm date mri mg evidence hospital report history examination noted
580-589	Nephritis, nephrotic syndrome, and nephrosis
	renal clip reason left chest failure line final catheter radiology report examination date medical underlying patient pleural condition dialysis hd
317-319	Mental retardation
	pt tube noted chest cc resp patient retardation thick secretions care cont plan trach abd hr ct neuro telemetry coarse
E910-915	Accidents caused by submersion, suffocation, and foreign bodies
	pt patient esophageal food namepattern care perforation pain impaction oral secretions aspiration time esophagus white wife hospital discharge doctor intubated

Table 1. (Continued)

950-957	Injury to nerves and spinal cord
	pt resp trach family care pain neuro injury plan vent intact hr gi skin thick secretions movement noted cord yellow
338-338	Pain
	pain pt mg tablet po sig doctor patient daily md blood ml iv namepattern discharge prn disp refills hr esophageal
905-909	Late effects of injuries, poisonings, toxic effects, and other external causes
	pt noted care pain intact wound skin cont vac family patient yellow vent drainage plan changed secretions abd resp remains
910-919	Superficial injury
	signal ml thoracic level ativan images foraminal spine fentanyl stenosis prn moderate pressure sbp seizures jump mild ligamentous abrasions ointment
E820-825	Motor vehicle non-traffic accidents
	pt trauma contrast family vehicle motor neuro mva head hr remains support intact skin ct trach mvc vent sp mri
890-897	Open wound of lower limb
	pt resp skin care support thick intact wound plan family secretions remains peep tube vent drainage hr stable cont bs
V20-29	Persons encountering health services in Circumstances related to Reproduction and development
	pt hr drainage abd continue continues support vent hct fluid family ativan mg husband cont skin cv resp white line
660-669	Complications occurring mainly in the course of labor and delivery
	pt patient blood pm uterine bleeding clip hct post reason vaginal date history abd artery units namepattern delivery discharge sp

4.2 Mortality Prediction

Table 2 shows mortality prediction results, with AUC, sensitivity, and specificity shown for baseline features, baseline + LDA topics, and baseline + Labeled-LDA topics with four label settings. The model using baseline + topic features from Labeled-LDA with 111 labels achieved an AUC of 0.835 for 30-day post-discharge mortality prediction. For 6-month post-discharge mortality prediction, baseline + Labeled-LDA with 111 and 94 labels performed closely with AUCs of 0.829. While both topic model derived features outperform the baseline in both 30-day and 6-month prediction model, baseline + LDA topics achieves somewhat higher AUCs than baseline + Labeled-LDA topics.

Table 2. Results of 30-day and 6-month mortality prediction.

Post-discharge Timeframe	Prediction Model	AUC	Sensitivity	Specificity
30-day	baseline	0.736	75.000	56.063
	baseline + LDA with 50 topics	0.860	86.607	70.569
	baseline + Labeled-LDA with 111 labels	0.835	85.714	63.204
	baseline + Labeled-LDA with 94 labels	0.834	86.607	63.652
	baseline + Labeled-LDA with 79 labels	0.832	86.607	63.596
	baseline + Labeled-LDA with 59 labels	0.831	89.286	59.563
6-month	baseline	0.776	71.831	70.343
	baseline + LDA with 50 topics	0.842	78.873	75.090
	baseline + Labeled-LDA with 111 labels	0.829	78.873	73.137
	baseline + Labeled-LDA with 94 labels	0.829	78.592	71.545
	baseline + Labeled-LDA with 79 labels	0.827	78.873	72.176
	baseline + Labeled-LDA with 59 labels	0.826	78.873	71.154

4.3 Topic Mortality

We applied probability of mortality defined by Marlin et al.[34] for each topic to investigate correlation between topics and mortality. Table 3 depicts ICD-9-CM codes with corresponding probability of mortality for 30-day and 6-month post-discharge periods. The results suggested "viral diseases accompanied by exanthem" (050-059), "dislocation" (830-839), and malignant neoplasm of "other and unspecified sites", "respiratory and intrathoracic organs" (190-199 and 160-165), and "other diseases of skin and subcutaneous tissue" (700-709) are the potentially important causes of death for 30-day post-discharge. For 6-month post-discharge mortality, malignant neoplasm of "other and unspecified sites", "respiratory and intrathoracic organs", "lymphatic and hematopoietic tissue" (190-199, 160-165, and 200-208), and "other diseases of skin and subcutaneous tissue" (700-709) were potentially important causes of death. On the other hand, one can see that "open wound of limb" (880-887 and 890-897), "superficial injury" (910-919), "complications of labor and delivery" (660-669), "complications mainly related to pregnancy" (640-649), "injury to blood vessels" (900-904), "homicide and injury purposely inflicted by other persons" (E960-969), and "suicide and self-inflicted injury" (E950-959) were ranked high in both 30-day and 6-month post-discharge survival.

Table 3. The probability of mortality for 111 topics for 30-day and 6-month; the top potential causes of death are highlighted in bold; frequency of ICD-9-CM codes

ICD-9-CM	Definition	30-day	6-month	Frequency
001-009	Intestinal infectious diseases	0.0607	0.1752	371
030-041	Other bacterial diseases	0.0569	0.1404	1875
042-044	Human immunodeficiency virus (HIV) infection	0.0336	0.1203	149
050-059	Viral diseases accompanied by exanthem	**0.1644**	0.1693	85
070-079	Other diseases due to viruses and chlamydiae	0.0467	0.1365	593
110-118	Mycoses	0.0707	0.1483	410
130-136	Other infectious and parasitic diseases	0.0367	0.0664	87
150-159	Malignant neoplasm of digestive organs and peritoneum	0.0271	0.1146	415
160-165	Malignant neoplasm of respiratory and intrathoracic organs	**0.1190**	**0.2723**	289
170-175	Malignant neoplasm of bone, connective tissue, skin, and breast	0.0072	0.1139	79
179-189	Malignant neoplasm of genitourinary organs	0.0355	0.0826	222
190-199	Malignant neoplasm of other and unspecified sites	**0.1134**	**0.3519**	768
200-208	Malignant neoplasm of lymphatic and hematopoietic tissue	0.0764	**0.2443**	292
210-229	Benign neoplasms	0.0095	0.0332	331
235-238	Neoplasms of uncertain behavior	0.0949	0.2122	152
240-246	Disorders of thyroid gland	0.0267	0.0939	1264
249-259	Diseases of other endocrine glands	0.0270	0.0840	4022
260-269	Nutritional deficiencies	0.0726	0.2033	369
270-279	Other metabolic and immunity disorders	0.0094	0.0356	6912
280-285	Anemia	0.0382	0.1156	3731
286-287	Coagulation/hemorrhagic	0.0843	0.1648	1170
288-289	Other	0.0491	0.1894	289
290-294	Organic psychotic conditions	0.0674	0.1359	1219
295-299	Other Disorders	0.0245	0.0450	494
300	Neurotic disorders	0.0021	0.0332	450
303-305	Psychoactive substance	0.0079	0.0295	1773
306-311	Other (primarily adult onset)	0.0182	0.0667	707
317-319	Mental retardation	0.0110	0.0259	71
320-327	Inflammatory diseases of the central nervous system	0.0277	0.0705	377
330-337	Hereditary and degenerative diseases of the central nervous system	0.0319	0.1157	600
338-338	Pain	0.0219	0.0400	60
340-349	Other disorders of the central nervous system	0.0208	0.0768	938
350-359	Disorders of the peripheral nervous system	0.0065	0.0620	611
360-379	Disorders of the eye and adnexa	0.0331	0.0848	617
380-389	Diseases of the ear and mastoid process	0.0501	0.1752	119
393-398	Chronic rheumatic heart disease	0.0156	0.0499	585
401-405	Hypertensive disease	0.0058	0.0200	7452
410-414	Ischemic heart disease	0.0184	0.0502	5416

Table 3. (Continued)

415-417	Diseases of pulmonary circulation	0.0609	0.1183	763
420-429	Other forms of heart disease	0.0394	0.1068	7165
430-438	Cerebrovascular disease	0.0358	0.0745	1562
440-448	Diseases of arteries, arterioles, and capillaries	0.0175	0.0620	1521
451-459	Diseases of veins and lymphatics, and other diseases of circulatory system	0.0624	0.1487	1880
460-466	Acute respiratory infections	0.0394	0.0570	121
470-478	Other diseases of the upper respiratory tract	0.0250	0.0514	215
480-488	Pneumonia and influenza	0.0439	0.1523	1889
490-496	Chronic obstructive pulmonary disease and allied conditions	0.0537	0.1452	2452
500-508	Pneumoconioses and other lung diseases due to external agents	0.0744	0.1845	1059
510-519	Other diseases of respiratory system	0.0725	0.1877	3904
520-529	Diseases of oral cavity, salivary glands, and jaws	0.0147	0.0504	152
530-537	Diseases of esophagus, stomach, and duodenum	0.0266	0.0747	2200
550-553	Hernia of abdominal cavity	0.0240	0.0486	314
555-558	Noninfectious enteritis and colitis	0.0221	0.1213	364
560-569	Other diseases of intestines and peritoneum	0.0475	0.1165	1353
570-579	Other diseases of digestive system	0.0567	0.1277	2025
580-589	Nephritis, nephrotic syndrome, and nephrosis	0.0771	0.1749	2824
590-599	Other diseases of urinary system	0.0716	0.1831	2401
600-608	Diseases of male genital organs	0.0225	0.0925	434
617-629	Other disorders of female genital tract	0.0016	0.0641	102
640-649	Complications mainly related to pregnancy	0.0000	0.0000	86
660-669	Complications occurring mainly in the course of labor and delivery	0.0000	0.0000	51
680-686	Infections of skin and subcutaneous tissue	0.0285	0.0988	466
690-698	Other inflammatory conditions of skin and subcutaneous tissue	0.0003	0.0644	294
700-709	Other diseases of skin and subcutaneous tissue	**0.1270**	**0.2938**	667
710-719	Arthropathies and related disorders	0.0279	0.0655	677
720-724	Dorsopathies	0.0072	0.0387	485
725-729	Rheumatism, excluding the back	0.0164	0.0655	387
730-739	Osteopathies, chondropathies, and acquired musculoskeletal deformities	0.0676	0.1775	771
745-747	Circulatory system	0.0000	0.0007	352
780-789	Symptoms	0.0318	0.1036	3482
790-796	Nonspecific abnormal findings	0.0372	0.1182	959
797-799	Ill-defined and unknown causes of morbidity and mortality	0.0953	0.1847	218
800-804	Fracture of skull	0.0191	0.0375	423
805-809	Fracture of neck and trunk	0.0126	0.0415	774
810-819	Fracture of upper limb	0.0074	0.0297	324
820-829	Fracture of lower limb	0.0307	0.0703	353
830-839	Dislocation	**0.1382**	0.1382	94
850-854	Intracranial injury, excluding those with skull fracture	0.0383	0.0779	627
860-869	Internal injury of thorax, abdomen, and pelvis	0.0106	0.0278	600
870-879	Open wound of head, neck, and trunk	0.0067	0.0181	406
880-887	Open wound of upper limb	0.0000	0.0000	106
890-897	Open wound of lower limb	0.0000	0.0000	56
900-904	Injury to blood vessels	0.0000	0.0000	107
905-909	Late effects of injuries, poisonings, toxic effects, and other external causes	0.0000	0.0350	59
910-919	Superficial injury	0.0000	0.0255	58
920-924	Contusion with intact skin surface	0.0000	0.0159	136
930-939	Effects of foreign body entering through Body orifice	0.0854	0.1509	112
950-957	Injury to nerves and spinal cord	0.0300	0.0300	64
958-959	Certain traumatic complications and unspecified injuries	0.0429	0.0429	106
960-979	Poisoning by drugs, medicinal and biological substances	0.0068	0.0224	281
990-995	Other and unspecified effects of external causes	0.0479	0.1591	924

Table 3. (Continued)

996-999	Complications of surgical and medical care, not elsewhere classified	0.0364	0.1201	3880
E810-819	Motor vehicle traffic accidents	0.0034	0.0092	608
E820-825	Motor vehicle non-traffic accidents	0.0000	0.1293	56
E849	Place of Occurrence	0.0133	0.0370	488
E850-858	Accidental poisoning by drugs, medicinal substances, and biologicals	0.0302	0.0345	102
E870-876	Misadventures to patients during surgical and medical care	0.0583	0.0591	78
E878-879	Surgical and medical procedures as the cause of abnormal reaction of patient or later complication, without mention of misadventure at the time of procedure	0.0222	0.0315	1080
E880-888	Accidental falls	0.0367	0.0862	871
E910-915	Accidents caused by submersion, suffocation, and foreign bodies	0.0674	0.1033	66
E916-928	Other accidents	0.0180	0.0333	135
E930-949	Drugs, medicinal and biological substances causing adverse effects in therapeutic use	0.0296	0.0770	807
E950-959	Suicide and self-inflicted injury	0.0000	0.0132	194
E960-969	Homicide and injury purposely inflicted by other persons	0.0000	0.0000	131
V07-09	Persons with need for isolation, Other potential health hazards and Prophylactic measures	0.0741	0.1728	332
V10-19	Persons with potential health hazards related to personal and family history	0.0343	0.1025	2603
V20-29	Persons encountering health services in Circumstances related to Reproduction and development	0.0000	0.0000	53
V40-49	Persons with a condition influencing their health status	0.0358	0.1020	2281
V50-59	Persons encountering health services for specific procedures and aftercare	0.0363	0.0997	1162
V60-69	Persons encountering health services in other circumstances	0.1328	0.1373	222
V70-82	Persons without reported diagnosis encountered during examination and investigation of individuals and populations	0.0000	0.0000	79

5. Discussion

Our results confirm previous findings that LDA-derived topic features provide a promising boost to mortality prediction[1-2]. Although the features derived from the "vanilla" LDA achieve slightly higher AUC than Labeled-LDA, the "vanilla" LDA topics require domain experts to interpret the topics and associate them with the underlying disease representation. At the same time, our proposal of using Labeled-LDA model with ICD-9-CM codes as labels suggest a feasible way to achieve direct interpretability of topic features. Specifically, the top words of a topic derived with Labeled-LDA tend to be strongly associated with the corresponding definition of ICD-9-CM code. Note that expert evaluation of topic quality is also made easy by virtue of associating topics with ICD-9 definitions.

Our *transparent predictive model* effectively provides the ability to tailor mortality prediction to the particular diagnosis, with the Labeled-LDA topic model supplying an association score for each ICD-9-CM code via topic proportions. Several cancers are notorious causes of death as can be seen in Table 3. Likewise, the largest proportion of our patient cohort are cardiac patients, and Table 3 shows low average probability of mortality for the corresponding topics. On the other hand, dislocation was surprisingly ranked high in topic mortality. We examined the patients with dislocation and found that more than half of the patients were over 50 years old. It might suggest the poor recovery from dislocation of the elderly which causing following complications after they were discharged from the hospital.

We expect that using different methods of label selection to supplement frequency thresholding we used in this work may lead to improved prediction for the labeled LDA model. This may entail, for example, selecting the ICD-9 codes from specific levels of the ICD-9 hierarchy. This can be seen as similar to the topic granularity experiments in which the number of topics is changed in the regular LDA model.

Interestingly, the Labeled-LDA topic model can potentially be used to uncover relations between different diagnostic labels by virtue of examining the associated terms. As an example, some of top words in the topic associated with "Other metabolic and immunity disorders" such as 'chest', 'artery', 'coronary', 'cabg', and etc. may reflect the relationship between cardiac and metabolic diseases as described in Alvarez et al.[35] and Naschitz et al.[36] This suggests that using labeled LDA models which factor in the label frequency and interdependence, such as the ones proposed by Rubin et al.[37] can potentially be used to explore the correlation between different labels.

6. Conclusions

We demonstrated the promising predictive power for 30-day and 6-month post-discharge mortality prediction using Labeled-LDA derived topic features. Because of the diversity and complexity of the diseases, our approach incorporated ICD-9-CM codes as knowledge input to guide topic model learning. Given an ICU record, the derived model could be used to determine the likelihood of post-discharge mortality and provide the physician with a justification for this assessment in the form of a combination of diagnostic codes associated with derived high-risk topics. In addition, ICD-9-CM topic features may be interpreted directly by healthcare professionals and patients for understanding the specific results of mortality prediction. In future work, different ICD-9-CM hierarchy and Labeled-LDA variants may be explored to improve the topic interpretability and prediction model.

7. Acknowledgements

References

1. Ghassemi M, Naumann T, Joshi R, Rumshisky A. Topic models for mortality modeling in intensive care units. In ICML Machine Learning for Clinical Data Analysis Workshop 2012.
2. Ghassemi M, Naumann T, Doshi-Velez F, et al. Unfolding physiological state: mortality modelling in intensive care units. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining 2014 Aug 24;(pp. 75-84). ACM.
3. Jo Y, Loghmanpour N, Rosé CP. Time series analysis of nursing notes for mortality prediction via a state transition topic model. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management 2015 Oct 17;(pp. 1171-1180). ACM.
4. Le Gall JR, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. Jama. 1993 Dec 22;270(24):2957-63.
5. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Critical care medicine. 1985 Oct 1;13(10):818-29.
6. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. Intensive care medicine. 1996 Jul 1;22(7):707-10.
7. Lehman LW, Saeed M, Long WJ, Lee J, Mark RG. Risk stratification of ICU patients using topic models inferred from unstructured progress notes. In AMIA 2012 Nov.
8. Khan MS, Maitree P, Radhika A. Evaluation and comparison of the three scoring systems at 24 and 48 h of admission for prediction of mortality in an Indian ICU: a prospective cohort study. Ain-Shams Journal of Anaesthesiology. 2015;8(3):294.
9. Moon BH, Park SK, Jang DK, Jang KS, Kim JT, Han YM. Use of APACHE II and SAPS II to predict mortality for hemorrhagic and ischemic stroke patients. Journal of Clinical Neuroscience. 2015;22(1):111-115.
10. Geerse DA, Span LF, Pinto-Sietsma SJ, van Mook WN. Prognosis of patients with haematological malignancies admitted to the intensive care unit: Sequential Organ Failure Assessment (SOFA) trend is a powerful predictor of mortality. European journal of internal medicine. 2011;22(1):57-61.
11. Saeed M, Villarroel M, Reisner AT, et al. Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): a public-access intensive care unit database. Critical care medicine. 2011 May;39(5):952.
12. Arnold CW, Oh A, Chen S, Speier W. Evaluating topic model interpretability from a primary care physician perspective. Computer methods and programs in biomedicine. 2015 Oct 30.
13. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In Proceedings of the AMIA Symposium 2001;(p.17).
14. Friedman C. A broad-coverage natural language processing system. In Proceedings of the AMIA Symposium 2000;(p.270).
15. Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. Journal of the American Medical Informatics Association. 2010;17(5):507-513.
16. Boag W, Wacome K, Tristan Naumann MS, Rumshisky A. CliNER: a lightweight tool for clinical named entity recognition.

17. Gobbel GT, Reeves R, Jayaramaraja S, et al. Development and evaluation of RapTAT: a machine learning system for concept mapping of phrases from medical narratives. Journal of biomedical informatics. 2014 Apr 30;48:54-65.
18. Salmasian H, Freedberg DE, Friedman C. Deriving comorbidities from medical records using natural language processing. Journal of the American Medical Informatics Association. 2013 Dec 1;20(e2):239-42.
19. Cobb R, Puri S, Wang DZ, Baslanti T, Bihorac A. Knowledge extraction and outcome prediction using medical notes. In ICML workshop on Role of Machine Learning in Transforming Healthcare. 2013
20. Alaniz Macedo A, Pollettini JT, Munson EV. A chronic illness system using biomedical knowledge sources and relevance feedback. In Computer-Based Medical Systems (CBMS), IEEE 28th International Symposium. 2015 June;(pp. 244-249).
21. Karystianis G, Buchan I, Nenadic G. Mining characteristics of epidemiological studies from Medline: a case study in obesity. J. Biomedical Semantics. 2014;5:22.
22. Hanauer DA, Saeed M, Zheng K, et al. Applying MetaMap to Medline for identifying novel associations in a large clinical dataset: a feasibility analysis. Journal of the American Medical Informatics Association. 2014;21(5): 925-937.
23. Li Y, Salmasian H, Vilar S, Chase H, Friedman C, Wei Y. A method for controlling complex confounding effects in the detection of adverse drug reactions using electronic health records. Journal of the American Medical Informatics Association. 2014;21(2):308-314.
24. Nori N, Kashima H, Yamashita K, Ikai H, Imanaka Y. Simultaneous modeling of multiple diseases for mortality prediction in acute hospital care. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2015 Aug 10;(pp. 855-864). ACM.
25. Makar M, Ghassemi M, Cutler DM, Obermeyer Z. Short-term mortality prediction for elderly patients using Medicare claims data. International Journal of Machine Learning and Computing. 2015 Jun 1;5(3):192.
26. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. the Journal of machine Learning research. 2003 Mar 1;3:993-1022.
27. Teh YW, Jordan MI., Beal MJ, and Blei DM. Hierarchical dirichlet processes. Journal of the American Statistical Association, 2006;101(476):1566-1581.
28. Ramage D, Hall D, Nallapati R, Manning CD. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 2009 Aug 6;(pp. 248-256). Association for Computational Linguistics.
29. Salton G, McGill MJ. Introduction to modern information retrieval. 1986.
30. Andrzejewski D, Zhu X, Craven M, Recht B. A framework for incorporating general domain knowledge into latent dirichlet allocation using first-order logic. In IJCAI Proceedings-International Joint Conference on Artificial Intelligence. 2011 Jul;22(1):1171.
31. Andrzejewski D, Zhu X, Craven, M. Incorporating domain knowledge into topic modeling via dirichlet forest priors. In Proceedings of the 26th Annual International Conference on Machine Learning. 2009 June;(pp. 25-32). ACM.
32. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Medical care. 1998 Jan 1;36(1):8-27.
33. Chang CC, Lin CJ. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST). 2011 Apr 1;2(3):27.
34. Marlin BM, Kale DC, Khemani RG, Wetzel RC. Unsupervised pattern discovery in electronic health care data using probabilistic clustering models. In Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium 2012 Jan 28;(pp. 389-398). ACM.
35. Alvarez AM, Mukherjee D. Liver abnormalities in cardiac diseases and heart failure. Int J Angiol. 2011 Sep;20(3):135-42.
36. Naschitz JE, Slobodin G, Lewis RJ, Zuckerman E, Yeshurun D. Heart diseases affecting the liver and liver diseases affecting the heart. American heart journal. 2000 Jul 31;140(1):111-20.
37. Rubin TN, Chambers A, Smyth P, Steyvers M. Statistical topic models for multi-label document classification. Machine learning. 2012;88(1-2): 157-208.