

# Risk Stratification of ICU Patients Using Topic Models Inferred from Unstructured Progress Notes

Li-wei Lehman<sup>1</sup>, PhD, Mohammed Saeed<sup>2</sup>, MD, PhD, William Long<sup>3</sup>, PhD,  
Joon Lee<sup>1</sup>, PhD, Roger Mark<sup>1</sup>, MD, PhD

<sup>1</sup>Harvard-MIT Health Sciences and Technology, <sup>2</sup> University of Michigan, Ann Arbor,  
<sup>3</sup>Computer Science and Artificial Intelligence Laboratory, MIT

## Abstract

We propose a novel approach for ICU patient risk stratification by combining the learned “topic” structure of clinical concepts (represented by UMLS codes) extracted from the unstructured nursing notes with physiologic data (from SAPS-I) for hospital mortality prediction. We used Hierarchical Dirichlet Processes (HDP), a non-parametric topic modeling technique, to automatically discover “topics” as shared groups of co-occurring UMLS clinical concepts. We evaluated the potential utility of the inferred topic structure in predicting hospital mortality using the nursing notes of 14,739 adult ICU patients (mortality 14.6%) from the MIMIC II database. Our results indicate that learned topic structure from the first 24-hour ICU nursing notes significantly improved the performance of the SAPS-I algorithm for hospital mortality prediction. The AUC for predicting hospital mortality from the first 24 hours of physiologic data and nursing text notes was 0.82. Using the physiologic data alone with the SAPS-I algorithm, an AUC of 0.72 was achieved. Thus, the clinical topics that were extracted and used to augment the SAPS-I algorithm significantly improved the performance of the baseline algorithm.

## 1. INTRODUCTION

ICU acuity metrics are routinely utilized to quantitatively characterize the severity of illness of ICU patient populations, and are applied for mortality prediction and benchmarking ICU performance across different institutions. Acuity metrics typically utilize structured admission data (from the first 24 hours) including: laboratory results, vital signs, and known chronic diseases diagnoses. We hypothesized that there is rich clinical information within unstructured nursing progress notes from the first 24 hours that are predictive of hospital outcomes and severity of illness.

To utilize this additional source of information, we applied topic models [1–3] to automatically learn “topic” structure of clinical concepts, represented by Unified Medical Language System (UMLS) codes, extracted from unstructured nursing text during the first 24-hours of patients’ ICU stays. Our approach combined the learned topic structure of clinical concepts from the nursing notes with physiologic data (from SAPS-I) to predict adverse patient outcomes. We used Hierarchical Dirichlet Processes (HDP) [3], a non-parametric topic modeling technique, to automatically discover “topics” as shared groups of co-occurring UMLS clinical concepts. We evaluated the potential utility of the inferred topic structure in predicting hospital mortality using the nursing notes of 14,739 adult ICU patients from the MIMIC II database [4].

Topic modeling is a Bayesian learning approach originally developed for information retrieval [1]. In topic models, documents are represented by un-ordered sets of words. By discovering patterns of word use from documents that exhibit similar patterns, words are grouped into thematically coherent structure, called topics. A *topic* is thus a set of words that tend to co-occur, and represents word co-occurrence patterns that are shared across multiple documents in the corpus. HDP automatically discovers topics that are shared across documents in the corpus. The number of topics is assumed to be unknown a priori, and is inferred from the data. A document can be associated with multiple topics; the topic proportion of each document is defined by the proportion of words assigned to each topic. Topic models have been successfully applied to various domains, including gene expression analysis [5], clinical data analysis [6], image analysis [7], and computer vision [8].

To apply topic modeling in unstructured clinical text, we represented patients using un-ordered sets of UMLS codes extracted from the nursing notes. For each ICU patient, clinical concepts from the first 24-hour ICU nursing notes were combined to form a single “document.” A “word” in our case is thus a UMLS code, representing either a disease, symptom, medication, procedure, or finding from the patients’ nursing notes. Discovered topics thus represent sets of UMLS codes that tend to co-occur in a collection of nursing notes.

\* Manuscript received March 15, 2012.

## 2. METHODOLOGY

### 2.1. Patient Selection

Nursing notes from the first 24 hour ICU stay of each adult patient in MIMIC II (version 2.5) were extracted. Patients whose SAPS-1 [9] score could not be determined due to missing data were excluded, as were patients whose lengths-of-stay were less than 24 hours. For patients with multiple ICU stays or multiple hospital stays, notes from the first ICU stay of the first hospital stay were used.

### 2.2. Data Preparation

UMLS clinical concepts were extracted from unstructured nursing text using a previously described [10] natural language processing (NLP) technique. We used a simple parser [10] to identify medical concepts in the nursing notes. The parser uses the UMLS as its medical dictionary using the SNOMED-CT and RXNORM concepts primarily. The concepts have types and we only used concepts whose types could be considered as diseases, symptoms, findings, medications, and procedures. UMLS codes that occur less than 5 times in the entire corpus were eliminated from the study.

The parser divides the text into phrases using punctuation and conjunctions. Within the phrases it performs a longest first search by normalizing the phrase and looking it up in the UMLS normalized string table. If no concept is found, it removes the last word from the phrase and tries until it finds a UMLS concept or runs out of words. It then moves forward a word and tries again. In this way it finds the maximum length concepts, some of which may be overlapping. Thus it finds the most specific concept in the text (“anterior myocardial infarction”, not just “myocardial infarction”).

Because of ambiguities in the mapping of phrases to concepts, especially among abbreviations, we have generated an additional dictionary for providing preferred meanings for some the more common ambiguities. However, since new versions of the UMLS add new concepts and abbreviations, this is an ongoing process. If a phrase has multiple mappings, the program favors diseases or other desired types. For example, “RA” would be considered rheumatoid arthritis rather than radial artery. Thus, it favors sensitivity over specificity.

The parser also associates other concepts as modifiers to the desired concepts. Of importance here are the negative modifiers, such as “no” or “ruled out”. These negative phrases are from NEGEX [11]. If a concept has a negative associated with it, it is removed from consideration.

### 2.3. HDP Settings

For HDP parameter settings, we used the same notations as in [3]. A two-level hierarchical Dirichlet process implementation [12] was used to build our topic models. We used a symmetric Dirichlet distribution with parameters of 0.2 for the prior  $H$  over topic distributions. The concentration parameters were given vague gamma priors,  $\gamma$  Gamma(1, 0.1) and  $\alpha_0$  Gamma(1, 1). Results presented are output of the model after 1000 iterations of Gibbs sampling. Performance reported was averaged across ten different topic models generated by ten different runs of the Gibbs sampling algorithm.

### 2.4. Evaluation Methods and Statistical Analysis

Multivariate logistic regression was performed to find the association between each topic (proportion) variable and hospital mortality. Specifically, the topic proportion of each document (defined as the proportion of words assigned to each topic) was used as input to multivariate logistic regression for hospital mortality prediction. SAPS-I was used as a control variable. For each topic variable, we computed its p value and odds ratio (OR, with 95% confidence interval) after adjusting for SAPS-I.

Hospital mortality prediction performance was evaluated using 10-fold cross validation with learned topic structure as well as SAPS-I as inputs to multivariate logistic regression. Forward search was used for feature selection. Continuous variables were expressed as mean with standard deviation or median with interquartile range where applicable, and compared with t test or the Wilcoxon rank sum test. Categorical variables were described as proportions and compared with the Chi-square test. Two-sided p values less than 0.05 were considered statistically significant.

	Sensitivity	Specificity	PPV	NPV	AUC
SAPS-I	0.64 ( $\pm$ 0.01 )	0.67 ( $\pm$ 0.01 )	0.25 ( $\pm$ 0.003 )	0.92 ( $\pm$ 0.001 )	0.72 ( $\pm$ 0.0003 )
Topic Prop	0.72 ( $\pm$ 0.01 )	0.70 ( $\pm$ 0.02 )	0.29 ( $\pm$ 0.01 )	0.94 ( $\pm$ 0.001 )	0.78 ( $\pm$ 0.01 )
Topic Prop + SAPS-I	0.76 ( $\pm$ 0.02 )	0.72 ( $\pm$ 0.01 )	0.32 ( $\pm$ 0.01 )	0.95 ( $\pm$ 0.003 )	0.82 ( $\pm$ 0.003 )

Table 1. Prediction performance for hospital mortality. SAPS-I performance were averaged across ten iterations of 10-fold cross validations. Performance reported for topic models were 10-fold cross validated AUCs averaged across ten different topic models generated by ten runs of the Gibbs sampling algorithm.

### 3. RESULTS

#### 3.1. Characteristics of the Cohort

Nursing notes (during the 1st 24 hrs of patient’s ICU stay) for 14,739 adult patients from the MIMIC II database were included in this study. The hospital mortality of these patients was 14.6% (2,154 expired at end of their hospital stays). The average ( $\pm$  STD) age of the patient population was 64.50 ( $\pm$  20.77), with median and interquartile range of 65.90 (51.98, 77.80). 57% of the patient population was male. The number of unique UMLS terms in this corpus was 3,678. The total number of UMLS terms across the corpus was 602,665 (average 40.9 UMLS terms per patient).

#### 3.2. Prediction Performance

We ran HDP 10 times using parameter settings described in the Methodology section. The number of topics generated (from 10 different runs of the Gibbs sampling algorithm) ranged from 39 to 44. Table 1 details the prediction performance averaged across ten different topic models generated by ten different runs of the Gibbs sampling algorithm. Using SAPS-I alone, the average AUC for hospital mortality prediction was 0.72 ( $\pm$  0.0003). Using topic proportions, the average AUC for hospital mortality prediction was 0.78 ( $\pm$  0.01). Using topic proportions in combination with the SAPS-I variable, the average AUC for hospital mortality prediction was 0.82 ( $\pm$  0.003).

#### 3.3. Clinical Relevance of Learned Topics

The model used to present our main results in Table 2 contained a total of 42 topics; the topics from this model were typical across multiple runs of the algorithm. Multivariate logistic regression was performed to find the association between each of the 42 topic (proportion) variables and hospital mortality. A total of 22 topic variables had p values  $\leq$  0.05 after adjusting for SAPS-I. We present ten of those topics in Table 2. These ten topics were reviewed by a clinician, and each topic was assigned a clinical label that best described the “topic” based on a review of the top 10 words in each topic. We report the top ten words associated with each topic, as well as the clinician assigned labels. For each topic, we report its odds ratio (OR, with 95% confidence interval) for hospital mortality. Odds ratios were defined per 10% increase in topic proportion after adjusting for SAPS-I.

Our results based on multivariate logistic regression indicate significant associations between several learned topics and hospital mortality (see Table 2). In particular, topics related to “hemorrhagic stroke”, “liver failure”, “sepsis”, “heart failure”, “on ventilator”, “cancer”, “hypotension” were significantly associated with increased hospital mortality (with odds ratios greater than one), indicating that increasing proportions of these topics were associated with an increased chance of hospital mortality.

To define a patient subgroup for each topic, we included patients with topic proportion  $\geq$  10% for that topic to form a subgroup. For each patient subgroup, we reported hospital mortality, age (median and interquartile range), as well as the top five most common ICD9 codes for patients in that subgroup. It should be noted that patient subgroups defined based on topic proportions in Table 2 are not mutually exclusive.

In comparison to the baseline (mortality 14.6%), all ten patient subgroups in Table 2 have statistically significant different mortality rates than the baseline population (Chi-square p values  $\leq$  0.001). Seven subgroups have significantly higher mortality than the baseline (Chi-square p values  $\leq$  0.001); three subgroups have significantly lower mortality than the baseline (Chi-square p values  $\leq$  0.001). In general, the top 10 UMLS “words” listed in each topic were clinically coherent within a topic, and corresponded well with the ICD-9 codes assigned to the patient subgroup.

Topic Label by Clinician	Top UMLS Descriptions		Odds Ratio (95% CI) N (Mortality%), Age	ICD-9	
Post-cardiac surgery	Extubation	CABG	0.70 (0.68 0.73) 3419(6%) 68 (58, 77)	414.01	CORON ATHEROSCLER NATIVE
	Insulin	CPAP		427.31	ATRIAL FIBRILLATION
	Drainage procedure	Coughing		411.1	INTERMED CORONARY SYND
	Propofol	Morphine		428.0	CONGESTIVE HEART FAILURE
	Nitroglycerin	Surgical Incisions		424.1	AORTIC VALVE DISORDER
Hemorrhagic stroke, Brain bleed	ICP	Coughing	1.33 (1.28 1.38) 658 (42%) 64 (45, 78)	431	INTRACEREBRAL HEMORRHAGE
	Mannitol	Hemorrhage		430	SUBARACHNOID HEMORRHAGE
	Suction drainage	Drainage procedure		331.4	OBSTRUCTIV HYDROCEPHALUS
	Unresponsive	Propofol		518.81	RESPIRATORY FAILURE
	Gagging	Ventriculostomy		427.31	ATRIAL FIBRILLATION
Liver failure, Cirrhosis	Lactulose	Octreotide	1.34 (1.27 1.40) 562 (33%) 54 (47, 66)	571.2	ALCOHOL CIRRHOSIS LIVER
	Icterus	Cirrhosis of liver		570	ACUTE NECROSIS OF LIVER
	Hemorrhage	Apyrexial		571.5	CIRRHOSIS OF LIVER NOS
	Ascites	Ethanol		572.2	HEPATIC COMA
	TIPS	Male individual		456.20	BLEED ESOPH VAR OTH DIS
Sepsis, Pneumonia	Vancomycin	Apyrexial	1.20 (1.17 1.24) 1488 (30%) 77 (64, 85)	518.81	RESPIRATORY FAILURE
	DNR	Oxygen		038.9	SEPTICEMIA NOS
	Tube	Fibrillation-atrial		507.0	FOOD/VOMIT PNEUMONITIS
	Rehab	Coughing		428.0	CONGESTIVE HEART FAILURE
	Suction drainage	Lethargy		427.31	ATRIAL FIBRILLATION
Heart failure	Fibrillation-atrial	Oxygen	1.16 (1.12 1.20) 2321 (21%) 77 (66, 84)	428.0	CONGESTIVE HEART FAILURE
	Congestive heart failure	Hypertensive disease		427.31	ATRIAL FIBRILLATION
	Heparin	Apyrexial		518.81	RESPIRATORY FAILURE
	Dyspnea	Myocardial infarction		410.71	AMI, SUBENDOCARD INFARCT
	Rales	Amiodarone		584.9	ACUTE RENAL FAILURE NOS
On ventilator	PEEP	Sedation Procedure	1.11 (1.08 1.14) 2113 (30%) 65 (51, 77)	518.81	RESPIRATORY FAILURE
	Sedated state	Respiratory therapy		038.9	SEPTICEMIA NOS
	Suction drainage	Insulin		428.0	CONGESTIVE HEART FAILURE
	Fentanyl	Intubation		414.01	CORON ATHEROSCLER NATIVE
	Propofol	Acidosis		507.0	FOOD/VOMIT PNEUMONITIS
Cancer	Secondaries	CA - Breast cancer	1.37 (1.26 1.49) 415 (27%) 64 (54, 75)	518.81	RESPIRATORY FAILURE
	Chemotherapy regimen	Hemoptysis		198.3	SEC MAL NEO BRAIN/SPINE
	Lung cancer	Coughing		198.5	SECONDARY MALIG NEO BONE
	Neoplasms	Stent, device		486	PNEUMONIA, ORGANISM NOS
	Dyspnea	Hypersensitivity		197.0	SECONDARY MALIG NEO LUNG
Hypotension	Dopamine	IVF	1.22 (1.14 1.31) 821 (28%) 74 (62, 82)	428.0	CONGESTIVE HEART FAILURE
	Dopa	Dobutamine		518.81	RESPIRATORY FAILURE
	Hypotension	Catheterization		414.01	CORON ATHEROSCLER NATIVE
	Antihypertensive agent	Bradycardia		410.71	AMI, SUBENDOCARD INFARCT
	Atropine	Vomiting		584.9	ACUTE RENAL FAILURE NOS
Intoxication, ETOH withdrawal	The psyche	DTs	0.75 (0.66 0.84) 775 (6%) 47 (37, 58)	518.81	RESPIRATORY FAILURE
	Ethanol	Employed		507.0	FOOD/VOMIT PNEUMONITIS
	Toxic effect	General drug type		780.39	OTHER CONVULSIONS
	IVF	OD		291.81	ALCOHOL WITHDRAWAL
	Male individual	Alcohol abuse		584.9	ACUTE RENAL FAILURE NOS
Trauma	Cervical collar	Employed	0.92 (0.88 0.96) 775 (6%) 52 (34, 74)	860.0	TRAUM PNEUMOTHORAX-CLOSE
	Trauma	TLS		518.5	POST TRAUM PULM INSUFFIC
	Injury	Contusions		427.31	ATRIAL FIBRILLATION
	Abrasion	Male individual		861.21	LUNG CONTUSION-CLOSED
	Coughing	Film		305.00	ALCOHOL ABUSE-UNSPEC

Table 2. Association between topic proportions and hospital mortality. Odds ratios for hospital mortality are defined per 10% increase in topic proportions. Adjusted (for SAPS-1) p-values for all topics shown are  $\leq 0.001$ . Age shown is median with interquartile range. To define a patient subgroup for each topic, we included patients with topic proportion  $\geq 10\%$  for that topic to form a subgroup. N = number of patients in each sub-group. Hospital mortality of the baseline population is 14%. The most common five ICD-9 codes (top-3 ranked for each patient) in each subgroup are shown. Acronyms: Transjugular intrahepatic portosystemic shunt [procedure] (TIPS), Aortocoronary bypass for heart revascularization (CABG), Positive end expiratory pressure (PEEP), Chronic obstructive airway disease (COPD).

## 4. DISCUSSION

Measures of severity of disease such as SAPS [9, 13], APACHE [14] and newer scores such as those developed by Hug et al. [15], provide estimates for patient mortality, assisting the physician in prioritizing patients. Our approach complements existing works in acuity metrics by combining automatically learned structure from the nursing notes with traditional structured physiologic data from an electronic medical record (EMR) for patient risk stratification.

Our approach automatically identifies “topics” as sets of clinically coherent groups of co-occurring diseases, symptoms, and findings documented in the first 24-hour nursing notes of the study ICU patient population. Topic learning was done in a completely un-supervised manner; no prior medical knowledge of disease associations were used. Discovered association patterns among clinical concepts represent the shared characteristics of the patient population in the study. Results from multivariate analysis indicated that topic proportions generated by HDP contained predictive information for hospital mortality after adjusting for severity of illness (represented by SAPS-I). The AUC for predicting hospital mortality from the first 24 hours of physiologic data and nursing text notes was 0.82. Using the physiologic data alone with the well-known SAPS-I algorithm, an AUC of only 0.72 was generated. Thus, the clinical topics that were extracted and used to augment the SAPS-I algorithm significantly improved the performance of the baseline algorithm. To our knowledge, this study is among the first to apply Hierarchical Dirichlet Processes to progress notes for ICU patient risk stratification, and demonstrated that unstructured nursing text notes are enriched with clinically meaningful information that can be automatically extracted and utilized in clinical decision-support tools.

Prior work by Hug et al. [15] developed a real-time ICU acuity score based on several hundred variables and reported an AUC of 88% - 89% for hospital mortality prediction using the MIMIC II database. Their study performed an extensive search through several hundred physiological and clinical variables extracted or derived from structured data to achieve high prediction performance. In contrast, our approach focused on using latent information inferred from unstructured nursing notes for patient risk stratification. Future work is required to investigate whether the topic structure inferred from nursing notes can be used to enhance the prediction performance of newer acuity-scores such as those reported in [15].

Other previous works have focused on using NLP techniques from discharge summaries to code patient outcomes [16, 17]. In [18], Saria et al. demonstrated that significant performance improvement can be achieved by integrating structured data (such as medication, treatment and lab results) with information extracted from the discharge summaries to code patient outcomes. Others have proposed automated methods to find disease-finding associations [19–21]. However, they did not examine utilities of their findings in the context of hospital mortality prediction. In [22], Cohen et al. used clustering analysis to identify clinically relevant patients states after trauma; their analysis focused on physiological data, instead of data from clinical text.

This study has several limitations that should be noted. First, the algorithm was evaluated only on the MIMIC-II database collected from one academic medical center. Ideally, an automated acuity score should be evaluated with an ICU database representative of a diverse patient population from different medical centers. Additionally, SAPS-I instead of SAPS-II scores were used as a baseline algorithm, since not all variables in SAPS-II were readily available for all patients in the MIMIC II database. Prior work by Hug et al. [15], however, has computed a pseudo-SAPS II score for 1,954 MIMIC II patients, and reported an AUC of 80.9% for hospital mortality based on day 1 ICU stays. The prediction performance of our approach (by combining topic structure with SAPS-I) is thus comparable to the pseudo-SAPS II score reported in [15].

Second, information extracted from the nursing notes may contain noise, and may not represent hard facts about states or treatments received by the patient. For example, the mentioning of “Propofol” in a patient’s nursing notes does not guarantee administration of Propofol to the patient. However, we believe that since we characterize each patient not by the appearance of a single clinical term, but rather by a distribution over a collection of related UMLS concepts (i.e. topic proportions), our approach has the potential to attenuate the effects of false positives from individual terms. Further research is required to investigate the effect of the NLP algorithm on the topics formed and the prediction performance.

Third, although probabilistic topic modeling is a powerful technique for data mining and exploratory data analysis, the model output is not definitive (i.e., different output may be generated from different runs of the Gibbs sampling algorithm). As such, the co-occurrence relationship (among clinical concepts) uncovered within each topic does not prove significant or causal clinical associations.

Nevertheless, our findings suggest that “topics” learned represent automatically discovered co-occurrence relationships that maybe useful for hypothesis generation and may provide valuable information in defining new feature sets for predicting patient outcomes. Further, we find that the topic models generated from the same HDP parameter settings produce topics that are generally stable and consistent across different runs, with small variations in the word orderings (among the top ranked words) for most topics. Variations among predictive performance of different models are small, with a standard deviation of 0.01 for AUCs from using ten different topic models.

## 5. CONCLUSIONS

In this paper, we used Hierarchical Dirichlet Processes (HDP) mixture models to automatically discover latent structure embedded in the nursing notes from the first 24 hours of patients' ICU stays. Patient cohorts defined based on the learned topics have statistically significantly different mortality rates than the baseline population. Further, the learned "topic" structure from the first 24 hour ICU nursing notes contains predictive information for hospital mortality, and may be used to augment the performance of existing acuity scores such as SAPS-I for patient risk stratification. Finally, "topics" learned may provide new insights for disease associations and motivate further studies to improve patient care.

There are several extensions of this work that we plan to explore, including the incorporation of different sources of data, knowledge of interventions, imaging and other diagnostic tests and procedures. We plan to evaluate the utility of our approach both in the context of decision support for ICU patient monitoring and in identifying high-risk patients for long-term patient care post hospital discharge.

## 6. ACKNOWLEDGMENTS

We would like to thank Dr. Mengling Feng for his technical advice, and the reviewers for their feedback. This work was funded in part by the National Institute of Biomedical Imaging and Bio-engineering and by the National Institute of General Medical Sciences, under NIH cooperative agreement U01-EB-008577 and NIH grant R01-EB001659.

## References

- [1] Blei D, Lafferty J. Topic Models. In A. Srivastava and M. Sahami, editors, *Text Mining: Classification, Clustering, and Applications*. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, 2009.
- [2] Blei D, Carin L, Dunson D. Probabilistic Topic Models. *IEEE Signal Processing Magazine*, 55-65, Nov. 2010.
- [3] Teh YW, Jordan MI, Beal MJ, Blei D. Hierarchical Dirichlet Processes. *J. Acoust. Soc. Amer.*, 101(476), 1566-1582, 2005.
- [4] Saeed M, Villarroel M, Reisner A, Clifford G et al. Multiparameter intelligent monitoring in intensive care II (MIMIC-II): A public-access ICU database. *Crit Care Med*, 39(5), 2011.
- [5] Gerber GK, Dowell RD, Jaakkola TS, and Gifford DK. Automated Discovery of Functional Generality of Human Gene Expression Programs. *PLOS Computational Biology*, 3(8), August 2007.
- [6] Saria S, Koller D, Penn A. Learning individual and population level traits from clinical temporal data. *Proc. Neural Information Processing Systems (NIPS), Predictive Models in Personalized Medicine workshop*, December 2010.
- [7] Barnard K, Duygulu P, de Freitas N, Forsyth D, Blei D, and Jordan MI. Matching words and pictures. *J. Mach. Learn. Res.*, vol. 3, 1107-1135, March 2003.
- [8] Fei-Fei L and Perona P. A Bayesian hierarchical model for learning natural scene categories. *Proc. IEEE Computer Vision and Pattern Recognition*, 524-531, 2005.
- [9] Le Gall JR, Loirat P, Alperovitch A, et al. A simplified acute physiology score for ICU patients. *Critical Care Medicine*, 12(11), 975-977, 1984.
- [10] Long W. Extracting Diagnoses from Discharge Summaries. *AMIA 2005 Symposium Proceedings*, 470-474, 2005.
- [11] Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform.* 2001 Oct; 34(5):301-10.
- [12] Wang C. <http://www.cs.princeton.edu/~chongw/>.
- [13] 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, 270(24):2957-2963.
- [14] Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006, 34(5):1297-1310.
- [15] Hug CW, Szolovits P. ICU acuity: real-time models versus daily models. *AMIA Annu Symp Proc* 2009, 260-264.
- [16] Solt I, Tikk D, Gl V, Kardkoves ZT. Semantic Classification of Diseases in Discharge Summaries Using a Context-aware Rule-based Classifier. *J Am Med Inform Assoc.* 2009.
- [17] Friedman C, Shagina L, Lussier Y, Hripesak G. Automated encoding of clinical documents based on natural language processing. *J Am Med Inform Assoc.* 2004.
- [18] Saria S, McElvain G, Rajani AK, Penn AA, Koller DL. Combining Structured and Free-text Data for Automatic Coding of Patient Outcomes, *AMIA Annu Symp Proc.* 2010, 712-716, 2012.
- [19] Hanauer DA, Rhodes DR, and Chinnaiyan AM. Exploring Clinical Associations Using 'Omics' Basic Enrichment Analyses, *PLOS ONE*, Vol. 4, April 2009.
- [20] Cao H, Markatou M, Melton GB, Chiang MF, and Hripesak G. Mining a clinical data warehouse to discover disease-finding associations using co-occurrence statistics. *Proc. AMIA 2005 Symposium*, 106-110, 2005.

- [21] Xing Z, Pei J. Exploring Disease Association from the NHANES Data: Data Mining, Pattern Summarization, and Visual Analytics, *IJDWM*, 6(3), 11-27, 2010.
- [22] Cohen MJ, Grossman AD, Morabito D, Knudson MM, Butte AJ, Manley GT. Identification of complex metabolic states in critically injured patients using bioinformatic cluster analysis. *Critical Care*, 14:R10, 2010.