MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190386

Extracting Alcohol and Substance Abuse Status from Clinical Notes: The Added Value of Nursing Data

Maxim Topaz [a,b], Ludmila Murga [c], Ofrit Bar-Bachar [c], Kenrick Cato [a], Sarah Collins [a,d]

[a] School of Nursing, Columbia University, New York City, NY, USA
[b] Data Science Institute, Columbia University, New York City, NY, USA
[c] Cheryl Spencer Department of Nursing, University of Haifa, Haifa, Israel
[d] Department of Biomedical Informatics, Columbia University, New York City, NY, USA

Abstract

We applied an open source natural language processing (NLP) system "NimbleMiner" to identify clinical notes with mentions of alcohol and substance abuse. NimbleMiner allows users to rapidly discover clinical vocabularies (using word embedding model) and then implement machine learning for text classification. We used a large inpatient dataset with over 50,000 intensive care unit admissions (MIMIC II). Clinical notes included physician-written discharge summaries (n = 51,201) and nursing notes (n = 412,343). We first used physician-written discharge summaries to train the system's algorithm and then added nursing notes to the physician-written discharge summaries and evaluated algorithms prediction accuracy. Adding nursing notes to the physician-written discharge summaries resulted in almost two-fold vocabulary expansion. NimbleMiner slightly outperformed other state-of-the-art NLP systems (average F-score = .84), while requiring significantly less time for the algorithms development.: Our findings underline the importance of nursing data for the analysis of electronic patient records.

Keywords:

Substance-Related disorders, Alcoholism, Nursing informatics.

Introduction

Over the past decade, widespread adoption of health information technology, such as the electronic health record, resulted in rapidly growing volumes of clinical data. Most of these data are stored in an unstructured format, such as text narratives, within electronic health records across diverse healthcare settings. The narrative data require innovative tools that help clinicians and researchers to extract meaning from large databases of text.

Natural language processing (NLP) is a set of powerful techniques that can help in processing and deriving meaning from clinical narratives. NLP has the potential to identify clinicians' detailed documentation of wound information, mentions of gender identity, documentation of poor self-management status and other concerns about a patient expressed as free text [1–5].

Although significant progress has been made within the medical domain, for nursing and allied health professions NLP development and implementation remains relatively challenging and scarce[6,7]. Some of the challenges include lack of large datasets of labeled clinical notes from nursing or allied health professions needed to create machine learning and text mining algorithms. On the other hand, rule-based NLP systems are hard to construct because large curated vocabularies of nursing or allied health professions-specific lexicons are rarely available. Moreover, the importance of nursing data and its contribution to general tasks such as text classification, remains generally under-studied. We found only one study that used information extracted from nursing notes as an indicator of out-of-hospital mortality[8]. Our main contribution is further assessment of the importance of nursing data.

This study focuses on an important domain-identification of alcohol and substance abuse from clinical data. Prior research suggests that alcohol and substance abuse information is often documented as free text[9]. However, previous NLP studies used physician notes to extract alcohol and substance abuse information[10,11]. In this study, we explore the potential added value of nursing data.

In this study we applied our open source NLP application called "NimbleMiner" [12,13] to generate an algorithm that can automatically identify clinical notes with mentions of alcohol and substance abuse. Our system allows users to rapidly specify clinical vocabularies for a certain domain, apply weakly supervised rapid labeling, and then implement machine learning for text classification. In this study, we first used physician-written discharge summaries to generate the NLP algorithm and evaluate the algorithm's prediction accuracy. We then added nursing notes to the physician-written discharge summaries and evaluated the algorithm's prediction accuracy. We compared the performance of the two NLP algorithms to understand whether adding nursing notes resulted in better prediction accuracy when identifying clinical notes with mentions of alcohol and substance abuse. We also compared our system's performance to other state-of-the art NLP systems that were applied on the same documents.

Methods

Dataset

This study used the large, publically available, de-identified dataset MIMIC II which is comprised of data from adult patients admitted to the intensive care units (ICUs) at the Beth Israel Deaconess Medical Center from 2001 to 2012. The data included over 50,000 hospital ICU admissions. The dataset contained several types of clinical notes, including physician-written discharge summaries (n = 51,201) and nursing notes (n

= 412,343). Nursing notes included nursing admission notes, daily progress and status update notes, case management notes, etc. This study received an Institutional Review Board approval from the University of Haifa, Israel.

NLP System Description

Our NLP system NimbleMiner is an open source system developed by our team[12,13]. User manual and download options can be accessed at: http://github.com/mtopaz/NimbleMiner. Other research or clinical teams can use the system under the GNU General Public License v3.0. NimbleMiner includes several methodological stages of clinical note processing that are briefly described below and presented in Figure 1.

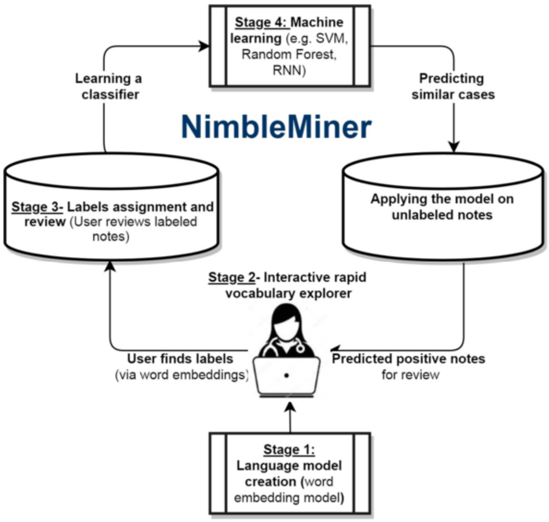

Figure 1– NimbleMiner system process stages.

Stage 1- Language Model Creation:

The user selects a large corpus of clinical notes and defines language model characteristics. We use a word embedding model for language model generation and users can change model settings based on their preferences.

Stage 2- Interactive Rapid Vocabulary

Explorer: The user enters a query term of interest, and the system returns a list of similar terms it identified as relevant. The list of suggested similar terms is based on the cosine term similarity metric extracted from the word embedding model. The user selects and saves the relevant terms. Negated or other irrelevant terms that are not selected by the user are also saved in the system for further tasks, such as negation detection. Figure 2 describes the steps of the vocabulary explorer stage.

Stage 3- Labels Assignment and Review:

The system uses the stage 2 discovered similar terms to assign labels to clinical notes (while excluding notes with negations and other irrelevant terms). Assigning a positive label means that a concept of interest is present in the clinical note. When needed, the user reviews and updates lists of similar terms and negated similar terms. The user reviews the clinical notes with assigned labels for accuracy. This weakly supervised rapid lableing approach is based on a postive labels learning framework validated in previous research [14,15].

Stage 4- Machine Learning:

The user choses a machine learning algorithm to be applied to create a predictive model. The model is then applied to predict what clinical notes might have the concept of interest. The user reviews the predicted notes and can go through stages 2-4 again to add new labels.

NLP System Settings

NimbleMiner's user interface is implemented in R statistical package. To create a word embedding model, we used a skip-gram model implementation called word2vec and phrase2vec in R[16,17]. Parameters of the word embedding model were held constant based on parameters suggested in other studies of word embedding[18]. Specifically, we used a model with window width size = 10, vector dimension = 100, minimum word count = 5, negative sample size = 5, and sub-sampling = 1e-3. For each similar term entered by the user, the system presented 50 potentially similar terms based on the cosine similarity. Our previous experiments [12] showed that the random forest algorithm outperforms other approaches (e.g., J48 Decision trees, Support Vector Machines), hence we used this algorithm in the machine learning stage of this study. The algorithm was trained using only discharge summaries, as was the case in other studies. The Random Forest algorithm was used with default settings (number of iterations = 100, minimum number of instances = 1, minimum variance for split = 1e-3, depth = unlimited).

NLP System Performance Evaluation

The system performance evaluation was performed based on the publically available gold-standard testing dataset generated by Gerhman et al. [19]. To create the gold standard dataset, Gerhman et al. annotated 1,610 discharge summaries for presence of several patient phenotypes, including alcohol and substance abuse. Each discharge summary was labeled by at least two experts in medicine and health informatics [19]. Full agreement about each of the labels was achieved on all the cases. Overall, the annotated dataset included 155 instances of substance abuse and 196 instances of alcohol abuse. For both domains, a relatively high inter-rater agreement was achieved (Cohen's Kappa inter-rater agreement = .86). We applied our NLP algorithms on this dataset to predict alcohol or substance abuse for each of the discharge summaries. We calculated precision (defined as the number of true positives out of the total number of predicted positives), recall (defined as the number of true positives out of the actual number of positives) and F-score (F1, weighted harmonic mean of the precision and recall) to evaluate the performance of our system.

Comparison with Other NLP Systems

NimbleMiner's performance was compared to the results reported by Gerhman et al. who developed a convolutional neural network (CNN) for text classification[19]. CNN is a subtype of machine learning models called neural networks. CNN is built of one or more convolutional layers and then followed by one or more fully connected layers, as in a standard multilayer neural network[20]. CNNs were found to outperform other machine learning methods in several classification tasks, such as image recognition[21].

In addition, NimbleMiner's performance was compared to another open-access rule-based NLP system called the clinical Text Analysis and Knowledge Extraction System (cTAKES)[22]. cTAKES was applied by Gerhman et al. to identify alcohol and substance abuse instances in each of the notes in the testing set[19]. To pre-process the clinical notes,

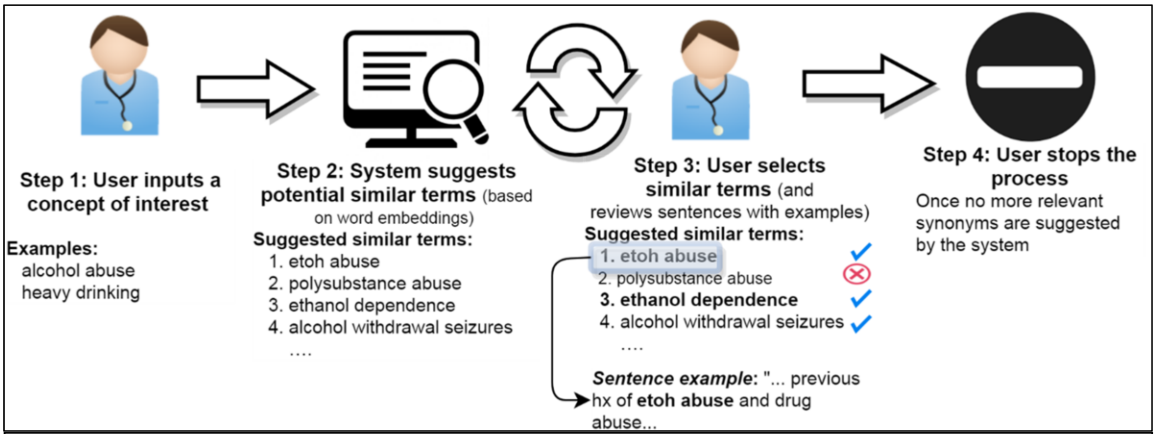

Figure 2– NimbleMiner Stage 2 -Interactive rapid vocabulary explorer process overview.

***Legend**: Interactive Rapid Vocabulary Explorer: The user enters a query term of interest (step 1), and the system returns a list of similar terms it identified as relevant (step 2). The list of suggested similar terms is based on the cosine term similarity metric extracted from the word embedding model. The user selects and saves the relevant terms while negated or other irrelevant terms that were not selected by the user are also saved in the system for further tasks (step 3). The system iteratively identifies new potential similar terms and presents them to the user for review (steps 2-3) and the process continues until there no new similar terms to suggest (step 4).*

cTAKES splits sentences and phrases into individual words (tokenization), normalizes them (e.g., removes plurals) and tags part-of-speech (e.g. noun, verb). Then, the named-entity recognition algorithm is implemented to detect named entities for which a concept unique identifier (CUI) exists in the Unified Medical Language System (UMLS).

Results

Applying NimbleMiner on discharge summaries alone resulted in identifying 73 alcohol abuse-related words and expressions (e.g., "heavy drinker" or "alcohol addiction"). Adding nursing notes resulted in identifying 71 additional unique words and expressions, resulting in a 97% lexicon expansion (total alcohol abuse expressions n = 144). Similarly, 76 substance abuse-related words and expressions (e.g., "illegal drug abuse" or "crack/cocaine abuse") were identified from the discharge summaries alone and 60 additional expressions were identified when nursing notes were added (total substance abuse expressions n = 136). Adding nursing notes enabled 79% lexicon expansion for the substance abuse domain. The lexicon discovery phase for each domain was conducted by one of the study co-authors and took about 4 hours in total to implement (2 hours for each domain) with discharge summaries only and about 6 hours when nursing notes were added (since more potentially relevant words needed to be reviewed). Two more hours were spend on system refinement (e.g., reviewing labeled notes and adding negations) and machine learning algorithm implementation for each domain. Overall, a maximum of 10 hours were spent on vocabulary development and algorithm implementation. Table 1 presents examples of alcohol and substance abuse words and expressions.

In addition, 1,736 terms for substance abuse and 2,255 terms for alcohol abuse were not selected by the user during the interactive rapid vocabulary explorer process (stage 2). These irrelevant terms included a diverse range of negations (e.g., "illicits none" or "not addicted to drugs"), family history expressions (e.g., "family history of alcoholism" or "father used cocaine") and other irrelevant terms that appear in the same context, for example other habits or related diseases (e.g., "tobacco smoker" or "hepatic cirrhosis"). Examples of negated or other irrelevant terms are shown in Table 1.

Table 1– Examples of alcohol and substance abuse expressions

Domain	Examples of relevant terms
Alcohol abuse	“heavy drinker” “alcohol addiction” “etoh [ethyl alcohol] abuse” “ciwa [The Clinical Institute Withdrawal Assessment for Alcohol scale] high” “korsakoff syndrome” [chronic memory disorder common in alcoholics] “alcoholic cirrosis [misspelling of cirrhosis]”
Substance abuse	“ho [history of] ivdu” “heroin withdrawal” “recent crack cocaine” “narcotic overdose” “iv heron [misspelling of heroin]” “cocaine in urine”
	Examples of negated or other irrelevant terms
Alcohol abuse	"drinks alcohol rarely" "occasionally drinks alcohol" "father was alcoholic" "family history of alcoholism" "alcohol abstinence" "social alcohol use"
Substance abuse	"ilicit drug use denies" "denies recreational drug use " "father used cocaine" "illicits none" "drugs none" "not addicted to drugs"

NLP System Performance Evaluation

NimbleMiner's predictive algorithm (Random Forest classifier) slightly outperformed other state-of-the-art NLP approaches in terms of the F-score. The best performing NimbleMiner algorithm was based on words and expressions learned from a corpus of discharge summaries and nursing notes compared to discharge summaries alone. See Table 2 for details on system performance measures

*Table 2– Comparison of system metrics across NLP systems**

Metric	NimbleMiner (discharge summaries+ nursing notes)	NimbleMiner (discharge summaries only)	Best performing Convolutional Neural Network	Text Analysis & Knowledge Extraction System (cTAKES)
		Alcohol abuse		
Precision	83	79	85	**88**
Recall	**84**	78	79	79
F1	**84**	78	81	83
		Substance abuse		
Precision	87	78	83	**93**
Recall	80	77	**83**	47
F1	**84**	77	83	62

*All systems were tested on discharge summaries only.

Discussion

This study applied a rapid NLP system called NimbleMiner to identify clinical notes with mentions of alcohol and substance abuse. Our approach showed promising results and it performed similarly or outperformed other systems developed for this domain[19]. For traditional machine learning and text mining, large datasets of labeled data are needed to train the system. For example, Gerhmann et al. annotated 1,610 discharge summaries and 70% of these notes were used for system training[19]. This approach required two or more clinicians to read and label the corpus of discharge summaries, resulting in significant time investment of at least 260 hours for data labeling (5 min per note * 1,610 discharge summaries * 2 clinicians= ~260 hours). On the other hand, the strength of rule-based systems lies in carefully curated and developed terminologies and rich hierarchical relationships between the concepts. However, rule-based systems like cTAKES require significant time and expertise investments when creating vocabularies and rules, defining negations, identifying family history, etc.

Our approach offers a hybrid solution where the human expert is interacting with the machine to rapidly create weakly supervised large labeled datasets for further machine learning. NimbleMiner also allows users to create large corpora of negated or other irrelevant terms (such as family history) that should be excluded from the labeled clinical notes processed through machine learning. Using the NimbleMiner, it took less time (4 hours) compared to other systems for algorithm development and implementation with comparable or better results. These findings suggest that for some text mining domains, such as identifying socio- behavioral determinants of health, NLP approaches can be implemented in a rapid, clinician-driven manner.

Our results also demonstrate the added value of extending data sources [23]. to include nursing data for vocabulary exploration. In our approach, adding nursing clinical notes resulted in an almost two-fold vocabulary expansion, leading to better classification performance. Other recent studies confirm our results about the importance of nursing data. For example, a NLP study conducted with the same dataset (MIMIC) has recently found that sentiment presented in nursing notes was significantly associated with a 30-day intensive care patient mortality[8]. Thus, the presence of negative-sentiment related expressions in nursing notes predicted a 30-day patient mortality. These findings underline the importance of nursing data for the analysis of electronic patient records. Other studies conducted in diverse clinical domains should consider using the valuable information from nursing notes to improve text and data mining algorithms.

Limitations

Our study has several limitations. First, our approach might not be applicable to solve more complex NLP challenges, such as word sense disambiguation or relation extraction. Comparing our results with rule-based systems like cTAKES would likely result in biased results in favor of more domain-specific systems like NimbleMiner. In addition, different NimbleMiner users can discover different vocabularies, which might influence the system's performance. Also, the vocabulary discovery phase was implemented by one co-investigator and adding more reviewers to this phase could have resulted in a different vocabulary for alcohol and substance abuse terms. A larger study with more domains is needed to validate the generalizability of our approach.

Conclusions

We applied our open-source NLP system NimbleMiner to conduct clinician-driven concept discovery in clinical narratives. Our results suggest that NimbleMiner can be applied to rapidly discover similar clinical terms and create large datasets of labeled clinical notes required for machine learning. Our system slightly outperformed other state-of-the art NLP systems while requiring significantly less time for the algorithm's development. We believe that NimbleMiner can be used by almost any clinician without special informatics training to create accurate NLP algorithms.

References

[1] S.A. Collins, and D.K. Vawdrey, Reading between the lines; of flow sheet data: nurses' optional documentation associated with cardiac arrest outcomes., *Appl. Nurs. Res.* **25** (2012) 251–7. doi:10.1016/j.apnr.2011.06.002.

[2] S.A. Collins, M. Fred, L. Wilcox, and D.K. Vawdrey, Workarounds used by nurses to overcome design constraints of electronic health records., *NI 2012 11th Int. Congr. Nurs. Informatics, June 23-27, 2012, Montr. Canada. Int. Congr. Nurs. Informatics (11th 2012 Montr. Quebec)*. **2012** (2012) 93.

[3] S.A. Collins, K. Cato, D. Albers, K. Scott, P.D. Stetson, S. Bakken, and D.K. Vawdrey, Relationship between nursing documentation and patients' mortality., *Am. J. Crit. Care*. **22** (2013) 306–13. doi:10.4037/ajcc2013426.

[4] K. Cato, J. Romano, R. Vanguri, and Ni. Tatonetti, The Use of Informatics to Reduce Disparities in Transgender Health, in: MIA 2017 Annu. Symp. - Sess. Details, Washington D.C., 2017.

[5] M. Topaz, K. Lai, D. Dowding, V.J. Lei, A. Zisberg, K.H. Bowles, and L. Zhou, Automated identification of wound information in clinical notes of patients with heart diseases: Developing and validating a natural language processing application, *Int. J. Nurs. Stud.* (2016). doi:10.1016/j.ijnurstu.2016.09.013.

[6] M. Topaz, and L. Pruinelli, Big Data and Nursing: Implications for the Future, in: Stud. Health Technol. Inform., 2017: pp. 165–171. doi:10.3233/978-1-61499-738-2-165.

[7] G. Gonzalez-Hernandez, A. Sarker, K. O'Connor, and G. Savova, Capturing the Patient's Perspective: a Review of Advances in Natural Language Processing of Health-Related Text, *Yearb. Med. Inform.* **26** (2017) 214–227. doi:10.15265/IY-2017-029.

[8] I.E.R. Waudby-Smith, N. Tran, J.A. Dubin, and J. Lee, Sentiment in nursing notes as an indicator of out-of-hospital mortality in intensive care patients, *PLoS One*. **13** (2018) e0198687. doi:10.1371/journal.pone.0198687.

[9] B. Tai, and A.T. McLellan, Integrating information on substance use disorders into electronic health record systems., *J. Subst. Abuse Treat.* **43** (2012) 12–9. doi:10.1016/j.jsat.2011.10.010.

[10] E. Chen, and M. Garcia-Webb, An analysis of free-text alcohol use documentation in the electronic health record: early findings and implications., *Appl. Clin. Inform.* **5** (2014) 402–15. doi:10.4338/ACI-2013-12-RA-0101.

[11] Y. Wang, E.S. Chen, S. Pakhomov, E. Arsoniadis, E.W. Carter, E. Lindemann, I.N. Sarkar, and G.B. Melton, Automated Extraction of Substance Use Information from Clinical Texts., *AMIA ... Annu. Symp. Proceedings. AMIA Symp.* **2015** (2015) 2121–30. http://www.ncbi.nlm.nih.gov/pubmed/26958312 (accessed March 18, 2019).

[12] M. Topaz, K. Gaddes, M. McDonald, and K. Bowles, Development and Validation of a Novel Rapid Clinical Text Mining Approach Based on Word Embeddings (NimbleMiner)., *Stud. Health Technol. Inform.* **244** (2017) 88.

[13] M. Topaz, L. Murga, K.M. Gaddis, M. V. McDonald, O. Bar-Bachar, Y. Goldberg, and K. Bowles, Mining fall-related information in clinical notes: comparison of rule-based and novel word embedding-based machine learning approaches, *J. Biomed. Inform.* (2019) 103103. doi:10.1016/J.JBI.2019.103103.

[14] Y. Halpern, S. Horng, Y. Choi, and D. Sontag, Electronic medical record phenotyping using the anchor and learn framework., *J. Am. Med. Inform. Assoc.* **23** (2016) 731–40. doi:10.1093/jamia/ocw011.

[15] C. Elkan, and K. Noto, Learning classifiers from only positive and unlabeled data, in: Proceeding 14th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. - KDD 08, 2008: p. 213. doi:10.1145/1401890.1401920.

[16] T. Mikolov, K. Chen, G. Corrado, and J. Dean, Efficient estimation of word representations in vector space, in: ICLR 2013, 2013: pp. 1–9. doi:10.1162/153244303322533223.

[17] R Core Team, R: A language and environment for statistical computing., (2013). http://www.r-project.org/. (accessed April 13, 2018).

[18] B. Chiu, G. Crichton, A. Korhonen, and S. Pyysalo, How to train good word embeddings for biomedical NLP, in: Proc. 15th Work. Biomed. Nat. Lang. Process., 2016: pp. 166–174. doi:10.18653/v1/W16-2922.

[19] S. Gehrmann, F. Dernoncourt, Y. Li, E.T. Carlson, J.T. Wu, J. Welt, J.F. Jr., E.T. Moseley, D.W. Grant, P.D. Tyler, and L.A. Celi, Comparing deep learning and concept extraction based methods for patient phenotyping from clinical narratives, *PLoS One*. **13** (2018) e0192360. doi:10.1371/JOURNAL.PONE.0192360.

[20] A. Ng, J. Ngiam, C.Y. Foo, Y. Mai, C. Suen, A. Coates, A. Maas, A. Hannun, B. Huval, T. Wang, and S. Tandon, Unsupervised Feature Learning and Deep Learning Tutorial, (n.d.).

[21] A.L. Beam, and I.S. Kohane, Big Data and Machine Learning in Health Care, *JAMA*. **319** (2018) 1317. doi:10.1001/jama.2017.18391.

[22] G.K. Savova, J.J. Masanz, P. V Ogren, J. Zheng, S. Sohn, K.C. Kipper-Schuler, and C.G. Chute, Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications., *J. Am. Med. Inform. Assoc.* **17** (2010) 507–13. doi:10.1136/jamia.2009.001560.

[23] P.F. Brennan, and S. Bakken, Nursing Needs Big Data and Big Data Needs Nursing, *J. Nurs. Scholarsh.* **47** (2015) 477–484. doi:10.1111/jnu.12159.

Address for Correspondence

Max Topaz, mt3315@cumc.columbia.edu.