NimbleMiner: an Open Source Nursing-sensitive Natural Language Processing System Based on Word Embedding

Maxim Topaz, PhD, RN[1,2], Ludmila Murga, PhD[1], Ofrit Bar-Bachar, MSc, PT[1], Margaret McDonald, MSW[3], Kathryn Bowles PhD, RN[3,4]

[1] The Cheryl Spencer Department of Nursing, Faculty of Social Welfare and Health Science, University of Haifa, Haifa, Israel; [2] Harvard Medical School & Brigham and Women's Hospital, Boston, USA; 3 The Visiting Nurse Service of New York; 4 School of Nursing, University of Pennsylvania, Philadelphia, USA.

Abstract

This study develops and evaluates an open source software (called NimbleMiner) that allows clinicians to interact with word embedding models (skip-gram models- word2vec) with a goal of creating lexicons of similar terms. As a case study, we implemented the system to identify similar terms for patient fall history from homecare visit notes (n= 1,149,586) extracted from a large U.S. homecare agency. We experimented with several parameters of word embedding models in order to identify the most time-effective and high quality model. Models with larger word window width sizes (n=10) that present users with about 50 top potentially similar terms for each (true) term validated by the user, were most effective. NimbleMiner can assist building a thorough vocabulary of fall history terms in about two hours. For domains like nursing, our approach could offer a valuable tool for rapid lexicon enrichment and discovery.

Introduction

Rapid and widespread adoption of health information technology in healthcare has created very large datasets of health data. Currently, up to 80% of these data are captured as unstructured narratives, for example clinical notes (e.g., discharge summaries, care coordination notes, radiology reports, etc.)[1]. Practicing clinicians and healthcare researchers are required to make use of these unstructured data in their everyday work. However, given the exponentially growing volume of unstructured data, extracting meaningful insights from these data has become very challenging. This is part of the reason why clinicians (e.g., physicians or nurses) spend up to half of their time on reading patients' charts and generating clinical documentation[2,3].

Over the past few decades, natural language processing (NLP) approaches have been increasingly applied to help extract meaningful insights from health narratives. Some health disciplines (mostly medicine) have seen significant advances in NLP systems development over the years[1,4] while for other disciplines, like nursing or other allied health professions, NLP has remained relatively new and underdeveloped[5]. In this manuscript, we use the latest advances in deep learning to start bridging the gap and create an open source, nursing-data sensitive, NLP system.

Related Work

This study focused on developing a novel approach to identify similar terms in clinical texts. Identifying and measuring semantic term similarity is one of the core NLP tasks[6–8]. Term similarity is defined as the likeness (in the shape or form) between two or more terms[6], for example, pairs of similar terms are "fall" and "patient collapsed". Identifying similar terms is a critical first step for many further NLP tasks, such as regular expression search, text mining, standard terminology development and maintenance, etc.

Traditionally, similar terms are identified using either knowledge-based approaches or distributional based measures. Knowledge-based approaches utilize human-created and curated knowledge sources, such as standardized terminologies[7]. For example, using the standardized health terminology called Unified Medical Language System (UMLS), we find that a concept "fall" (CUI C0085639) has mappings to several other concepts, like "falls" or "falling down"[9]. However, terminology-based approaches have poor practical applicability[10]. For example, standardized terminologies would not usually include term abbreviations or misspellings that commonly occur in real-world clinical narratives. In addition, for domains like nursing, there are relatively few well maintained standardized terminologies (with mappings to other terminologies) that can allow identification of diverse lists of similar terms. On the other hand,

similar terms can be identified with distributional based measures that use distribution of terms within a corpus to compute similarity. Distributional based measures are commonly based on the assumption that terms that appear in similar contexts are related[6]. These approaches are usually implemented using a specific corpus of health narratives, most commonly articles or article abstracts extracted from databases, such as PUBMED[8]. Although distributional based measures showed promising results in identifying similar concepts in the biomedical domain[6–8], we did not find any studies evaluating the application of these methods on nursing narratives.

Recently, several new distributional based approaches based on deep learning have emerged, for example the word embedding models. Unlike the traditional NLP language models where words are represented as discrete symbols, word embedding language models compute how often each word co-occurs with its neighboring words in a large text corpus, and then map these count-statistics down to a condensed vector for each word. The resulting language models are therefore sensitive to the context of each word or phrase and can be used to predict a word from its neighbors. In this study, we use a specific type of word embedding called skip-gram model[11]. Skip-gram models are particularly good for representing large text corpora and predicting context-words from target words. Conceptually, larger text corpora generate more robust skip-gram models.

Recently, several research groups have started applying word embedding models to measure term similarity[8,12,13]. However, the existing work has several limitations. First, the studies often used scientific articles (extracted from PUBMED or similar databases) to train word embedding models and little is known about the applicability of the method for clinical narratives taken from electronic health records. In addition, even when clinical narratives were used as a source for embedding, little was discovered about the extent to which word embedding-based similar terms were clinically relevant.

This study aimed to extend the current knowledge about the value of word embedding models for clinical narratives from electronic health records. We evaluated an approach and created a user interface for rapid and interactive similar concept extraction system based on word embedding. We offer our system as an open access system for free use (http://github.com/mtopaz/NimbleMiner). As a case study, we used a large database of homecare visit notes (mostly nursing notes) and applied our framework to identify similar concepts related to patient's fall history.

Methods

We first describe a case study we used to test our approach, then present an overview of our system (called NimbleMiner) and its user interface (UI) and finally present the system evaluation metrics.

Case study: identifying simclins for fall history from homecare clinical notes

In this case study, our goal was to identify similar terms indicating patient's fall history. We define fall as "An event which results in an individual coming to rest inadvertently on the ground or lower object"[14]. As a data source, this study used a large corpus of homecare visit notes (n= 1,149,586) for 89,459 patients treated by clinicians of one of the largest homecare agencies in the United States (located in New York, NY) during 2015. Clinical notes were completed by visiting homecare clinicians (e.g. nurses, physical/occupational therapists, social workers, etc.) using the homecare agency's electronic health record. This study examined the narrative part of the homecare visit description. Clinical notes ranged from lengthy admission notes (often written by a registered nurse) to shorter progress notes (e.g. physical therapy progress notes). The average note length was about 150 words.

Step-by-step system description

Preliminary step: Word embedding model creation and specifications.

Our UI provides users with several options for creating a word embedding model for further similar term discovery (Figure 1, Box 1, "Model building" tab). First, we ask the user to identify a .csv-type file that contains all the text data in one column (Figure 1, Box 2). Text data usually includes clinical notes of different lengths and various types. The intention is to present the system with as many clinical notes as possible (ideally >1,000,000 notes) to enable optimal word embedding model creation.

Then, we ask the user to identify word window width for model creation and provide a graphic example of different word window sizes (Figure 1, Box 3). The intuition here is that larger word windows would allow word embedding models to learn more about context of a specific word. On the other hand, very large word windows can potentially

decrease the model's ability to identify similar concepts. In this study, we experimented with different word window widths to identify optimal settings on our test data. To prepare clinical notes for the word embedding model training, we pre-processed the notes and removed punctuation and made all letters lower-cased. Additionally, we converted frequently co-occurring words in the clinical notes into phrases with lengths of up to four words (4-grams). This is a common process in NLP where sets of co-occurring words are combined into phrases, and we found that combinations of up to 4 words would cover most of the expressions in the case study explained later in details. For example, "pt_fell_yesterday" was a common 3-gram in our case study.

The other components of the word embedding models were held constant based on parameters suggested in other studies of word embedding[15]. Specifically, we used a skip-gram model (using Google's word2vec implementation in R) [11] with vector dimension =100, minimum word count = 5, negative sample size= 5 and sub-sampling= 1e-3.

We also ask users to indicate how many similar terms they want to see for each term they identified as relevant. For the purposes of our application, we refer to the similar terms as "simclins" (SIMilar CLINical terms). In tune with a concept called "anchor" suggested by Halpern et al. in a series of recent studies[16,17]. We define simclins as words or phrases that have high positive predictive value in identifying the concept of interest. In other words, if a simclin is present, then the patient should almost always have the condition or a problem we are aiming to find. For example, phrases like "pt collapsed" or "she fell down" are considered simclins for the presence of fall history.

In our approach, potential similar terms are identified automatically based on an attribute of word embedding model called cosine distance. Cosine distance is most commonly used in high-dimensional positive spaces. For example, in information retrieval, each term is notionally assigned a different dimension and a document is characterized by a vector where the value of each dimension corresponds to the number of times that the term appears in the document. Cosine distance ranges between 0 and 1, and words or phrases that appear in similar contexts have a higher cosine distance measure. Unrelated words and phrases would have lower cosine distance. Thus, sorting words or phrases on the highest cosine distance to each other, can potentially help users to identify simclins. In this study, we experimented with capping the potentially similar term lists at different lengths, presenting the users (researchers or clinicians) with the top 25, 50 or 75 most related words or phrases. The idea here is that presenting users with more potentially similar terms would result in longer processing times but might be necessary to identify more simclins. NimbleMiner provides users with an option to change the length of similar term lists (Figure 1, box 4).

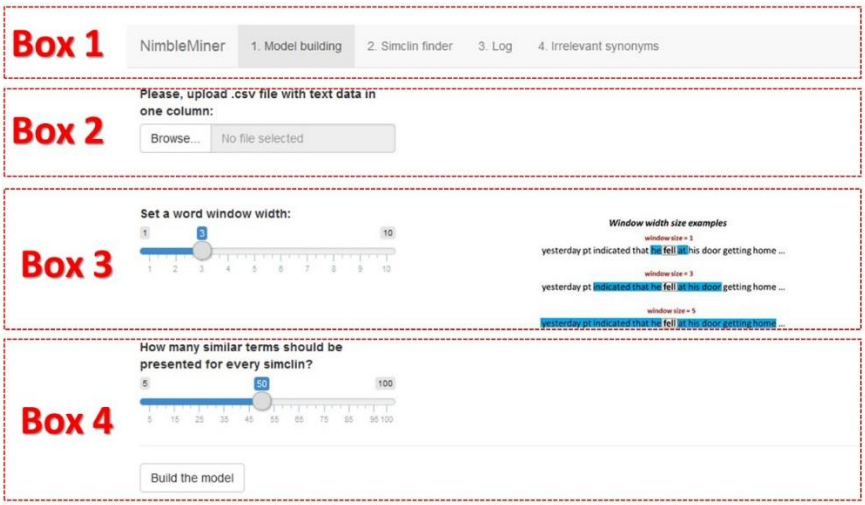

Figure 1. "Model building" tab in NimbleMiner (users can upload narrative files (Box 2), set a word window width (Box 3), and define how many similar terms are presented for every simclin (Box 4)).

Step 1: User inputs a concept of interest.

Figure 2 presents the general structure of our methodology. Step 1 starts with (Figure 2) the user who begins to input the desired simclins to find similar terms for. For example, in our case study of identifying fall history information in clinical notes, the user would start typing in words like "fall" or "fell", based on their clinical knowledge or an existing biomedical thesaurus. In NimbleMiner, simclins are typed in by the user on the "Simclin finder" page (Figure 3, box 1).

Once one or more simclins are added (Figure 3, box 2), they are presented in the simclins' list with examples from the clinical notes. Examples help the user to choose simclins accurately. NimbleMiner would then search for similar terms using the "Find similar terms for new simclins" button. The user can remove simclins from the list using the "Delete selected simclins" button.

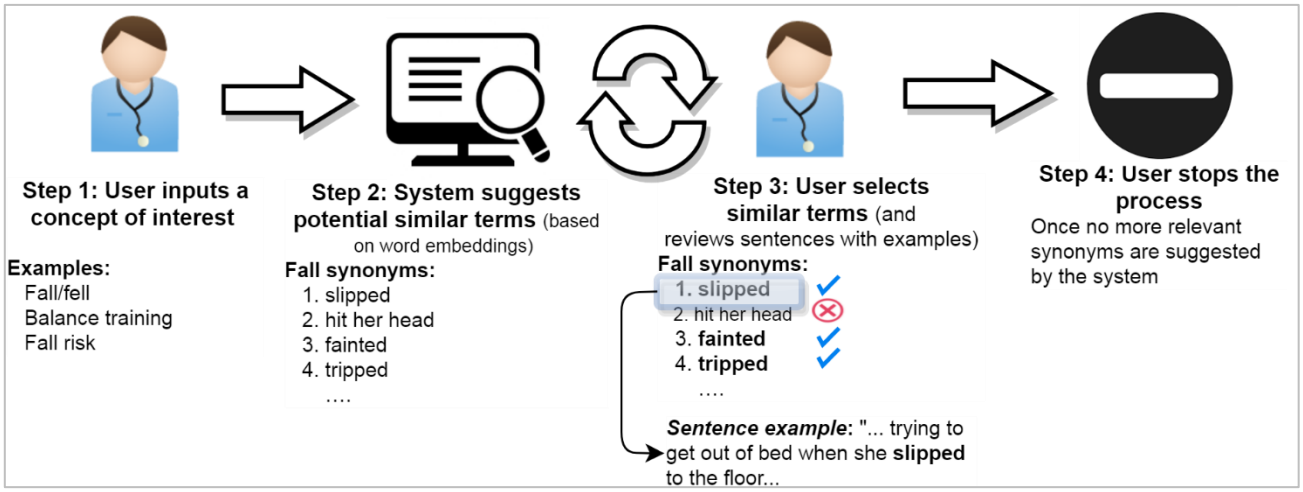

Figure 2. NimbleMiner methodology steps

For the purposes of the case study presented here, we only entered one word "fell" and all the other similar terms were discovered through model-suggested similar terms. All the case studies were completed by one expert in nursing and health informatics. We did this in order to enable better comparisons between the models.

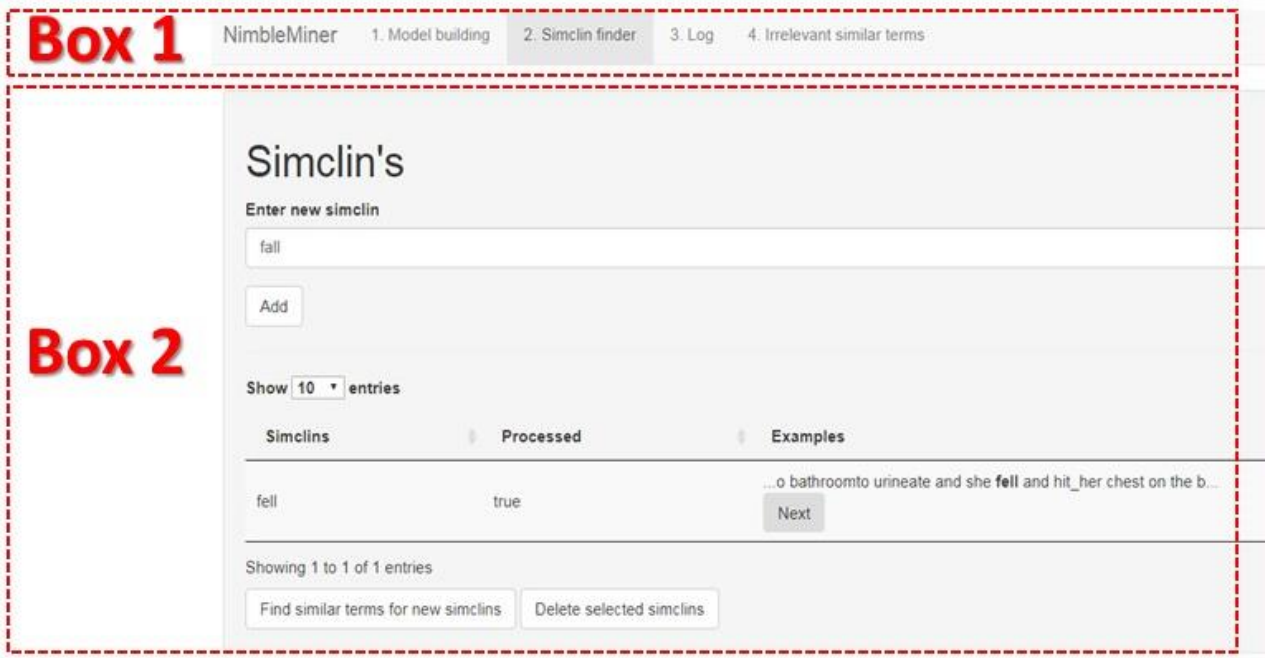

Figure 3. NimbleMiner's simclin finder tab (the user can specify simclins and review examples for the clinical notes (Box 2)).

Steps 2-3: System suggests potentially similar terms and the user selects (true) similar terms.

After the search for similar terms is completed, the user is presented with lists of most similar terms sorted by the cosine distance in the "New similar terms list" (Figure 4a). The user can simply click to select simclins out of the list of potentially similar terms. Examples from clinical notes are presented here again to help the user understand whether a certain term is a true simclin. For example, true simclins for fall history would include phrases like "fell down" or "had

fallen" while "almost fell" and "blacked out" are not true simclins as they indicate near fall or poorly specified loss of consciousness with no evidence of a fall. Once all relevant simclins are selected by the user, they are added to the list of simclins (using the "Save selected similar terms as simclins" button). Remaining similar terms that are not selected by the user are removed from the list and would not be shown again in further system iterations (using the "Clear all unselected clinical terms" button).

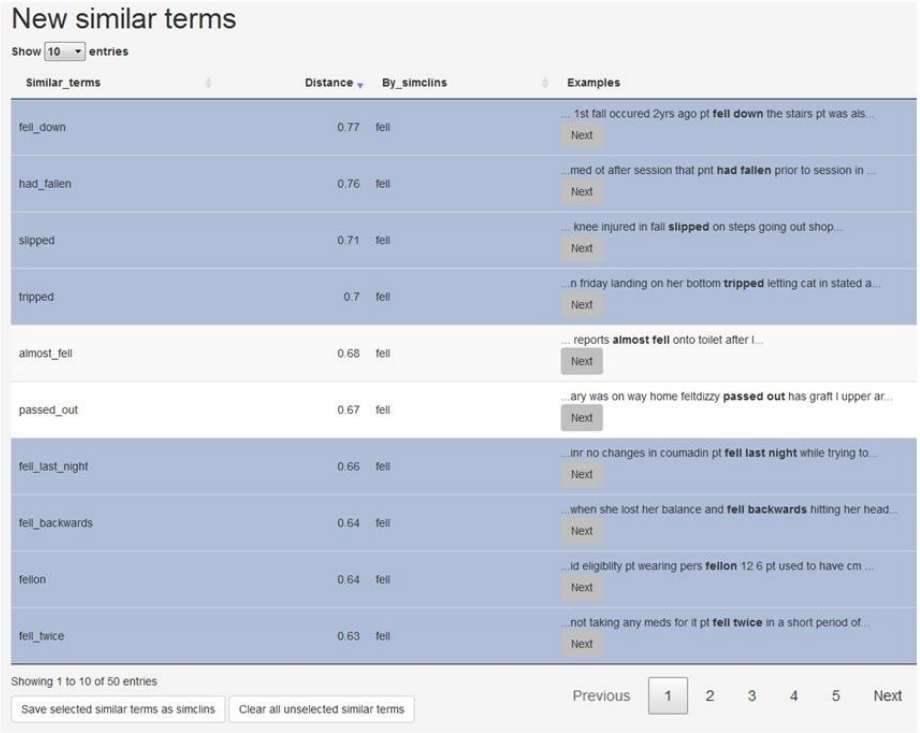

Figure 4a. NimbleMiner's potential similar term review window (the user can select simclins and review examples for the clinical notes).

The system would then iteratively search and present the user with lists of new similar terms for review. NimbleMiner would indicate when new potential similar terms are found on the lists of top similar terms. New potential similar terms are sorted by the highest value of the cosine distance. For example, Figure 4b shows that a suggested similar term (Figure 4b, box 1, "passed_out") is similar to many previously selected simclins (Figure 4b, box 2, "fainted, tripped, fell" etc.). In this situation, there is a higher likelihood that the new suggested similar term is indeed a simclin.

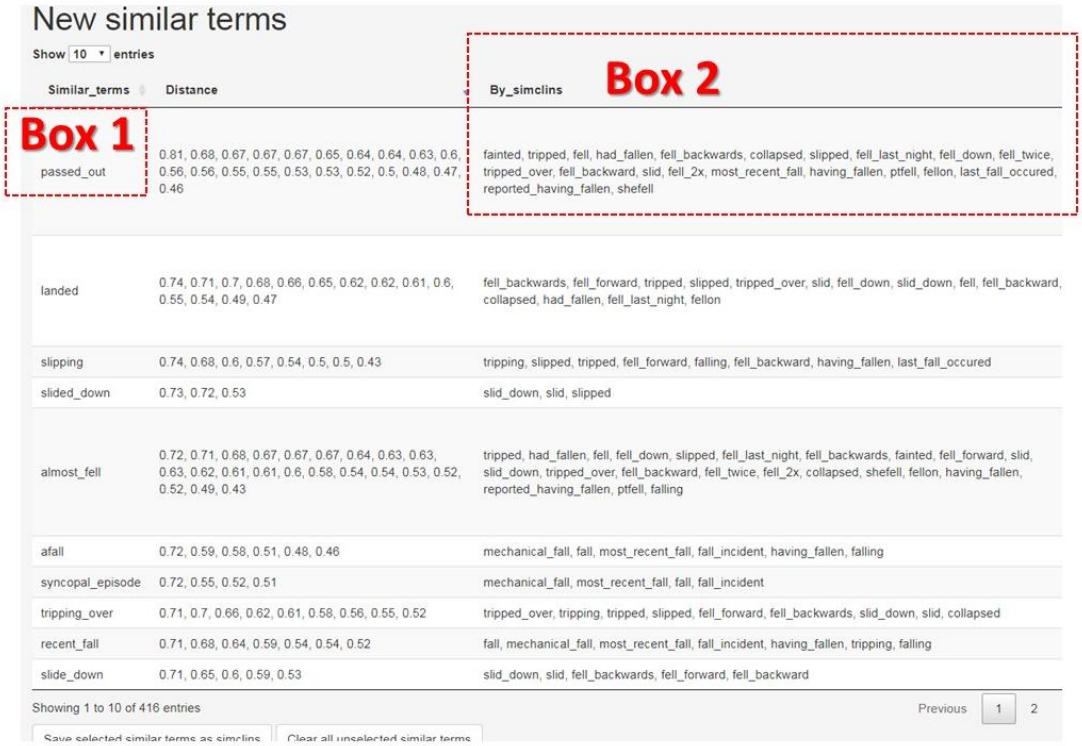

Figure 4b. NimbleMiner's potential similar term review window (users can select simclins and review examples for the clinical notes).

Step 4: User stops the process

Steps 2-3 are repeated until a.) the user cannot identify any additional simclins based on expertise or literature and b.) the user finished reviewing all system suggested potential new similar terms. The process stops at this point and simclins list can be exported as a .csv file. The user can easily review and reconsider terms that were previously removed in the "Irrelevant similar terms" tab and NimbleMiner also logs all the time and user action information in the exportable "Log" tab (Figure 1, box 1).

System evaluation metrics

This study tested four differently configured word embedding models with a goal of examining the differences in models' effectiveness, time spent working with each model and the resulting lexicon coverage. The models differed in terms of word window width (n= 3, 5, 7, 10) and lengths of top potentially similar term lists presented to the user (n= 25, 50, 75). In order to describe and compare models' effectiveness, we introduced several evaluation metrics.

- Simclin discovery effectiveness is a measure that reflects the percentage of true simclins out of all system-suggested potential similar terms and is specified as: Simclin discovery effectiveness (%)= (number of true simclins identified by the user / number of system suggested similar terms)*100. Higher simclin discovery effectiveness is a desired metric: for example, a 20% simclin discovery effectiveness would indicate that every fifth system-suggested word was chosen by the user as a simclin.
- We calculated how much time was needed on average to identify (true) simclins: Average simclin discovery time (seconds)= duration of the case study in seconds / number of true simclins identified by the user. In general, models with lower average simclin discovery times are preferred.
- We also compared the resulting lists of unique simclins between the models and evaluated the overlap between our simclin list and fall hierarchy (ID= 1912002) extracted from SNOMED-CT (US edition,[18]). Finally, for each model, we used the simclin list to conduct a regular expression search and presented the number of clinical notes. The idea is to evaluate the extent to which simclins identified by different models help discover "real world" cases.

Results

Word embedding models effectiveness and time comparisons

Table 1 summarizes the differences in evaluation metrics between the models. Both simclin discovery effectiveness and average simclin discovery times were highest for the model with highest word window width (n=10) when 25 potential similar terms were presented to the user (model J). The model was also second best in terms of the total time spent on the task (42 min). However, this model helped to uncover only 170 simclins vs. 236 simclins uncovered by the best model in terms of total simclins identified (model L). Model K was just 3 simclins short compared to the best model (model L) in terms of simclins found (233 vs. 236) while simclins discovery effectiveness and average simclin discovery time remained relatively low. We also present these results for the most effective models (J, K, and L) graphically in figure 5.

Table 1. Differences between word embedding models

Word Window Width	**3 words**			**5 words**			**7 words**			**10 words**		
Similar terms presented user (n)	**25**	**50**	**75**	**25**	**50**	**75**	**25**	**50**	**75**	**25**	**50**	**75**
Model	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Simclin discovery effectiveness (%)	6.6%	5.4%	3.8%	6.5%	6.3%	4.5%	6.1%	5.8%	4.5%	**7.7%**	6%	4.3%
Average discovery time (sec/simclin)	18.3	23.3	31.3	19.0	20.9	28.8	20.4	22.1	28.8	**14.8**	21.4	30.8
True simclins identified by the user (n)	131	139	161	164	215	217	221	214	233	170	233	**236**
System suggested similar terms (n)	1,970	2,576	4,256	2,512	3,436	4,843	3,614	3,721	5,199	2,201	3,900	**5,500**
Case study duration (min)	40	54	84	52	75	104	75	79	112	42	83	**121**
Number (%) of clinical notes with simclins* (total n= 1,149,586)	102,629 (8.9%)	103,098 (9%)	**107,660 (9.4%)**	60,358 (5.3%)	103,492 (9 %)	72,514 (6.3%)	103,918 (9%)	105,419 (9.2%)	104,453 (9.1%)	65,898 (5.7%)	105,923 (9.2%)	106,155 (9.2%)

* 51,050 (4.4%) clinical notes included words or expressions from the SNOMED-CT fall hierarchy (n=240 unique terms).

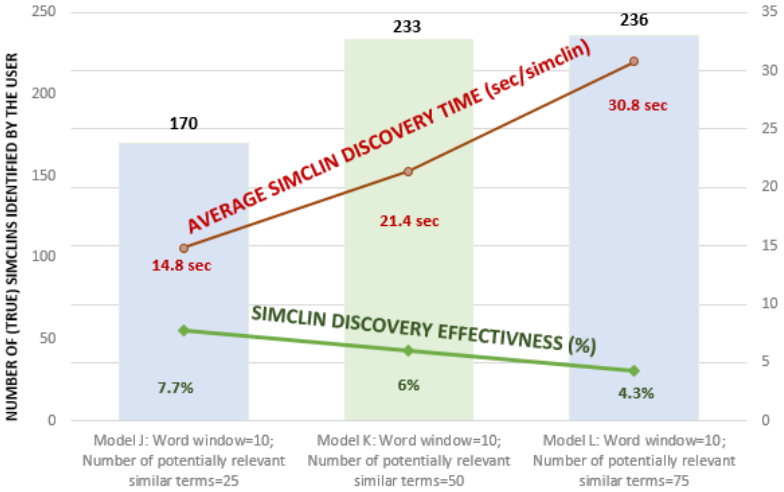

Figure 5. Number of true simclins identified by the user, simclin discovery effectiveness (%) and average simclin discovery time (sec/simclin).

The last row in Table 1 presents the number of unique clinical notes identified by regular expression search for all the simclins identified by each model. In addition, as a comparison, we searched for words and expressions of the "Fall" hierarchy in SNOMED-CT (240 unique terms). We found that using SMOMED-CT terms resulted in identifying 4.4% of the notes sample as potentially positive.

Simclins discovered by model C helped identify the largest number of clinical notes (9.4% of all the notes in the sample) while models H, K and L came very close (9.2%). Of note, this number of notes is preliminary and cannot be used as an estimate of the prevalence of falls in the sample. The current search was just for simclin regular expressions like “fall” and further post-processing is needed to exclude negated terms (e.g., “no falls”, “denies falls”, etc.), and other possible false positives.

When we compiled all the simclins produced by the different models into one list and excluded duplicates, we received 371 unique simclins. Overall, the list included 59% simclins (n= 220) that were variations of different expressions that included words like "fall" or "fell". These simclins included misspellings (e.g., "felll", "fals"), incorrectly written expressions (e.g., "felled") or expressions including these words (e.g., "mechanical fall", "ended up falling", or "fell off ladder"). The remainder of simclins (41%) included other words or phrases and their lexical variations, for example: "tripping over", "slipped off chair", "slided down", and "pt collapsed".

Lastly, when we examined the overlap between the full list of simclins and the 240 unique terms extracted from the "Fall" hierarchy in SNOMED-CT (e.g. "fall", "fall on concrete", "fall from bed"), we found only four shared terms.

Discussion

This paper is one of the first to report on development and evaluation of a user driven word embedding-based approach for similar term discovery from clinical notes. Similarly to others[8,12,13], our results show that word embedding models with larger window width sizes have more potential to help discover similar terms. We also found that when experts evaluate more potentially relevant similar terms, larger lists of true similar terms can be identified. On the other hand, using models that require users to review more potential similar terms results in increasing time needed to complete the task. For example, it took the user more than twice as long to implement our most time consuming model (model L, 121 min with 5,500 potential similar terms for user review) compared to the most effective model in terms of simclin discovery rate (model J, 42 min with 2,201 potential similar terms for user review). Further work is needed to experiment with using word embedding models with larger word width sizes and more words presented to users for review.

With some variation between the models, several models identified that up to 9% of the notes in the sample have simclins. Search for words and expressions from the SNOMED-CT resulted in identifying only half of the notes discovered by most of the other word embedding models. This is not surprising- standard terminologies provide standard terms for clinical expressions while clinicians have their own way of writing, which differs between settings and professions. Our study provides evidence that these profession or setting-specific lexicons can be learned relatively fast when clinicians are involved in the process. Our approach is also portable and could enable discovering new simclins in a streamlined fashion in different settings or among different health professions.

Based on the results of this study, we would advise on using the model that came very close to discovering as many simclins as the most comprehensive model but took a reasonably shorter time to implement (model K, 83 min with 3,900 potential similar terms for user review, finding 9.2% of clinical notes to contain simclins, depicted in the middle on Figure 5).

Most of the other studies that examined applying word embedding to find similar terms used terminology-based approaches. For example, a recent study by Percha et al. applied several word embedding models to identify radiology-related terms from radiology notes[13]. This study used an existing radiology terminology (called RadLex) that already had a list of similar terms (synonyms examined by experts) for many radiology related concepts. The study examined whether these terminology-based terms and their synonyms can be identified automatically using word embedding models trained on a large sample of radiology notes. However, the researchers found that similar terms could be discovered for only very small percentage of terminology terms (<10%), even when large lists of potentially similar terms were used. Our findings explain these results. Because word embedding models are designed to learn contextually similar words, they learn well form a corpus of clinical notes and help uncover lexical variants, misspellings, and other expressions related to a target term. In our study, only four terms were shared between an expert-validated subset of

simclins and an extensive standard health terminology SNOMED-CT. Our results indicate that word embedding models work best when human experts are involved in validating their output and our system (NimbleMiner) offers just that.

Further work

Our study describes and evaluates a relatively fast approach for discovering similar terms from clinical notes. We strongly believe that NimbleMiner can be easily applied, especially in domains like nursing, where little work has been done in the past in the field of large lexicon creation. In addition, we provide a tool that can be applied to easily create lexicons for concepts that were not previously structured and exist mostly in a narrative form, such as lack of social support, drug or alcohol abuse, poor living status and other socio-behavioral risk factors.

Importantly, identifying similar terms – or simclins- is just a first step in a thorough NLP process. Previously, we used simclins as a foundation for positive-labels-only[19] machine learning framework and showed promising results[20]. Using simclins for machine learning is just one example of their potential use. Other potential applications could include streamlined user-driven regular expression search, or standard terminology development and maintenance tasks. In addition, NimbleMiner is language agnostic. The system is using word embedding models that can be virtually trained for any language and we are currently evaluating NimbleMiner on clinical notes in Hebrew.

Limitations: Our study has several important limitations. First, we experimented with one data set and our results should be repeated and validated. In addition, we only experimented with a limited set of parameters that can be changed when word embedding models are trained, and further work should experiment with other model settings and clinical note corpus sizes. In addition, we did not evaluate our simclin vocabulary against vocabularies other than SNOMED-CT. We also do not know how many terms and phrases in our text corpus are describing fall history. Our combined models identified 371 unique simclins and it is challenging to estimate how many terms referring to falls were used in all the notes.

Conclusions

This study is one of the first to evaluate an approach where a human expert is interacting with word embedding models (skip-gram models) to evaluate and create lexicons of similar terms. Our results suggest that models with larger word window width sizes result in more similar terms identified by the user. The results also suggest that the most time-effective approaches could ask users to review lists of about 50 top potentially similar terms for each (true) term validated by the user. We show that in about two hours, a clinician who uses NimbleMiner could build a thorough vocabulary of fall history terms. Our system is intended to be used by clinicians with no prior knowledge in NLP or informatics, and we provide an open access to the system for further use and development. We strongly believe that for domains like nursing, where relatively little lexicon creation work has been done in the past, our approach could offer a valuable tool for lexicon enrichment and discovery. Further work should experiment with more word embedding model parameters, evaluate the system on languages other than English (as the approach is language agnostic), and identify the best machine learning environment that can benefit from fast, user-driven similar term discovery.

References

1. Meystre, S. M., Savova, G. K., Kipper-Schuler, K. C. & Hurdle, J. F. Extracting information from textual documents in the electronic health record: a review of recent research. Yearb. Med. Inform. 128–44 (2008).

2. Sinsky, C. et al. Allocation of Physician Time in Ambulatory Practice: A Time and Motion Study in 4 Specialties. Ann. Intern. Med. 165, 753–760 (2016).

3. Topaz, M. et al. Nurse Informaticians Report Low Satisfaction and Multi-level Concerns with Electronic Health Records: Results from an International Survey. AMIA ... Annu. Symp. proceedings. AMIA Symp. 2016, 2016–2025 (2016).

4. Jensen, P. B., Jensen, L. J. & Brunak, S. S. Mining electronic health records: towards better research applications and clinical care. Nat. Rev. Genet. 13, 395–405 (2012).

5. Topaz, M. & Pruinelli, L. Big Data and Nursing: Implications for the Future. in Studies in Health Technology and Informatics 232, 165–171 (2017).

6. Pedersen, T., Pakhomov, S. V. S., Patwardhan, S. & Chute, C. G. Measures of semantic similarity and relatedness in the biomedical domain. J. Biomed. Inform. 40, 288–299 (2007).

7. Garla, V. N. & Brandt, C. Semantic similarity in the biomedical domain: an evaluation across knowledge sources. BMC Bioinformatics 13, 261 (2012).

8. Zhu, Y., Yan, E. & Wang, F. Semantic relatedness and similarity of biomedical terms: examining the effects of recency, size, and section of biomedical publications on the performance of word2vec. BMC Med. Inform. Decis. Mak. 17, 95 (2017).

9. Kleinsorge, R., Tilley, C. & Willis, J. Unified Medical Language System (UMLS). Encyclopedia of Library and Information Science 369–378 (2002). doi:10.1002/9781118479612.ch16

10. Liu, Y., McInnes, B. T., Pedersen, T., Melton-Meaux, G. & Pakhomov, S. Semantic relatedness study using second order co-occurrence vectors computed from biomedical corpora, UMLS and WordNet. Proc. 2nd ACM SIGHIT Symp. Int. Heal. informatics - IHI '12 363 (2012). doi:10.1145/2110363.2110405

11. Mikolov, T., Corrado, G., Chen, K. & Dean, J. Efficient Estimation of Word Representations in Vector Space. Proc. Int. Conf. Learn. Represent. (ICLR 2013) 1–12 (2013). doi:10.1162/153244303322533223

12. Sabbir, A. K. M., Yepes, A. J. & Kavuluru, R. Knowledge-Based Biomedical Word Sense Disambiguation with Neural Concept Embeddings. Computation and Language (2016).

13. Percha, B. et al. Expanding a radiology lexicon using contextual patterns in radiology reports. J. Am. Med. Informatics Assoc. (2018). doi:10.1093/jamia/ocx152

14. Zecevic, A. A., Salmoni, A. W., Speechley, M. & Vandervoort, A. A. Defining a fall and reasons for falling: Comparisons among the views of seniors, health care providers, and the research literature. Gerontologist 46, 367–376 (2006).

15. Chiu, B., Crichton, G., Korhonen, A. & Pyysalo, S. How to Train good Word Embeddings for Biomedical NLP. in Proceedings of the 15th Workshop on Biomedical Natural Language Processing 166–174 (2016). doi:10.18653/v1/W16-2922

16. Halpern, Y., Choi, Y., Horng, S. & Sontag, D. Using Anchors to Estimate Clinical State without Labeled Data. AMIA ... Annu. Symp. proceedings. AMIA Symp. 2014, 606–15 (2014).

17. Halpern, Y., Horng, S., Choi, Y. & Sontag, D. Electronic medical record phenotyping using the anchor and learn framework. J. Am. Med. Inform. Assoc. 23, 731–40 (2016).

18. SNOMED. Snomed CT. U.S. National Library of Medicine (2016).

19. Elkan, C. & Noto, K. Learning classifiers from only positive and unlabeled data. in Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 08 213 (2008). doi:10.1145/1401890.1401920

20. Topaz, M., Gaddes, K., McDonald, M. & Bowles, K. Development and Validation of a Novel Rapid Clinical Text Mining Approach Based on Word Embeddings (NimbleMiner). Studies in Health Technology and Informatics 244, (2017).