

# Can a Convolutional Neural Network Support Auditing of NCI Thesaurus Neoplasm Concepts?

Hao Liu<sup>1</sup>, Ling Zheng<sup>2</sup>, Yehoshua Perl<sup>1</sup>, James Geller<sup>1</sup>, Gai Elhanan<sup>3</sup>

<sup>1</sup>Department of Computer Science  
New Jersey Institute of Technology,  
Newark, NJ USA  
{hl395, perl, geller}@njit.edu

<sup>2</sup>CSSE Department  
Monmouth University,  
West Long Branch, NJ USA  
zdzhenling@gmail.com

<sup>3</sup>Applied Innovation Center  
Desert Research Institute  
Reno, NV USA  
gelhanan@gmail.com

**Abstract**—We present a Machine Learning methodology using a Convolutional Neural Network to perform a specific case of an ontology Quality Assurance, namely discovery of missing *IS-A* relationships for Neoplasm concepts in the National Cancer Institute Thesaurus (NCIt). The training step checking all “uncles” of a concept is computationally intensive. To shorten the time and to improve the accuracy, we define a restricted methodology to check only uncles that are similar to each current concept. The restricted technique yields higher classification recall (compared to the unrestricted one) when testing against known errors found by domain experts who manually reviewed Neoplasm concepts in a prior study. The results are encouraging and provide impetus for further improvements to our technique.

**Keywords**—*CNN; Deep Learning; Neoplasm Hierarchy; National Cancer Institute Thesaurus; Quality Assurance; Abstraction Network; Machine Learning*

## I. INTRODUCTION

Ontologies play a major role in enabling precise communications and in support of healthcare applications, e.g. EHR systems. Many ontologies are large and complex. For example, the National Cancer Institute Thesaurus (NCIt) [1], serving cancer researchers inside and outside NIH, contains 135,243 concepts interrelated by 480,141 links in the April 2018 release. Due to their size and complexity, errors in ontologies are unavoidable. Users of ontologies such as the SNOMED ontology are concerned about errors [2]. Thus, quality assurance (QA) is essential in the lifecycle of ontologies [3]. For a summary of auditing (QA) techniques for ontologies and in particular for SNOMED and NCIt, see [4, 5].

However, QA resources for ontologies are typically scarce, while QA tasks are labor-intensive and time-consuming. Therefore, automated or semi-automated techniques that can either help in auditing an ontology or narrow down the places where to look for errors, are highly desired. Missing parent/child errors are particularly interesting to ontology curators, as the *IS-A* links are the backbone structure of an ontology, facilitating the inheritance of lateral relationships (called roles in NCIt).

Machine Learning (ML) has been proven successful in many fields, e.g., knowledge mining. ML was previously used in knowledge enrichment for ontologies [6-8]. However, can ML be used for quality assurance of ontologies in spite of the major difference between knowledge enrichment and quality

assurance? Knowledge enrichment mines external sources for new knowledge that does not exist in the ontology. However, QA discovers incorrect or missing knowledge. Consider missing *IS-A* relationships from an existing concept A to a concept B. If the concept B is already in the ontology, then adding an *IS-A* link between A and B is considered correcting an omission error. If concept B is not in the ontology and is added together with adding an *IS-A* from concept A to it, then this is knowledge enrichment. We note that curators of some ontologies, e.g., NCIt, are less interested in knowledge enrichment, unless required by users, than in quality assurance. In this paper, we attempt to use ML to address the task of detecting missing *IS-A* links between two existing concepts. This task is more challenging than knowledge enrichment, since it requires a judgement that concept A is a specification of concept B. For knowledge enrichment we only recognize that a concept is missing in the ontology and then insert it into the proper place.

In an unpublished study, we trained a Convolutional Neural Network (CNN) deep learning model to insert new concepts into the SNOMED CT ontology, i.e., an enrichment problem. In the present work, we train a CNN deep learning model to find missing parent/child errors in the Neoplasm subhierarchy of NCIt. The vector representations of concepts are obtained from an unsupervised neural network language model. The model is evaluated by its classification recall on an unseen dataset. We check the model’s classification recall by testing against 18 missing parent/child errors found by domain experts in a prior study [9]. Due to the size of the Neoplasm subhierarchy, the application of the training methodology is computation-intensive and time consuming.

In previous research we have introduced Abstraction Networks (AbNs) [10, 11]. An AbN provides a compact summarization and visual simplification of an ontology. The SABOC (Structural Analysis of Biomedical Ontologies Center) team at NJIT has demonstrated that Abstraction Networks are an effective tool to support quality assurance of ontologies [9, 12]. An *area taxonomy* [3], a type of Abstraction Network, is composed of meta-concepts called *areas*, connected by child-of links. An *area* (see Background section) represents a group of concepts with the same structure.

To accelerate the processing and improve recall, we modify the CNN methodology to limit its consideration, for each

concept, to the similar concepts of its area (in our formal sense of area). The modified, restricted methodology achieves 0.81 recall on the unseen testing data. It performs 50% better than the unrestricted methodology on the 18 known errors in terms of recall. The results for detecting missing *IS-A* links are not yet strong enough. However, the performance in recognition of known errors is encouraging and supports further improvement of our methodology in respect to CNNs and the use of AbNs.

## II. BACKGROUND

### A. Doc2vec

Numeric representation of variable-length texts, ranging from sentences to documents is a challenging task. Doc2vec, or Paragraph Vectors [13], an extension of word2vec (word embedding) [14], maps variable-length texts to fixed-length vectors. It is an unsupervised framework that learns continuous distributed vector representations from unlabeled text data of a paragraph/document, while preserving the inter-relationships of the text in the numeric format. In such vector representations, similar pieces of text are close to each other in Euclidean or cosine distance in lower dimensional vector spaces. The Doc2vec inherits the semantics of the words in the context, and takes the word order into consideration when constructing the representation. The latter advantage is important to our problem, as word order in our setup carries the concepts' topological/hierarchical order in the ontology. This is useful information for feature learning. To the best of our knowledge, this is the first study to derive vector representations for biomedical ontology classes via Doc2vec.

### B. CNN

Convolutional Neural Networks (CNN), initially invented for image recognition, have been widely used for various applications, including vision, speech recognition, and language translation. CNN models have also been successfully applied to solve various Natural Language Processing (NLP) problems such as search query retrieval [15], semantic parsing [16] and sentence modeling [17]. CNN utilizes convolving filters to automatically learn and extract local features from various layers, regardless of the input size. This makes CNN a very powerful tool for classification or prediction tasks, e.g. text classification [18] and relation extraction [19], even if the data or features have not been manually labeled for learning purposes. To the best of our knowledge, this is the first effort to adopt the CNN model for ontology quality assurance.

### C. Neoplasms of NCI

NCIt is published monthly by the National Cancer Institute (NCI) in OWL and flat file formats. It is a cancer reference terminology that is widely used. It covers cancer-related terminology in various fields, e.g., clinical care and translational and basic research. Concepts are linked to other concepts (parent concepts) in the same hierarchy by *IS-A* relationships. A concept may have multiple parent concepts. The semantics of concepts are defined by lateral relationships

(called "roles"), e.g. *Disease Has Associated Anatomic Site*.

Due to the NCI's cancer focus, the Neoplasm subhierarchy of NCIt is composed of 9,955 concepts. It is a core component of the *Disease, Disorder or Finding*, the largest hierarchy with 35,081 concepts, and is modeled with more detail, compared to non-neoplasm concepts in the hierarchy.

### D. Areas and Area Taxonomy

An area taxonomy [3, 20, 21] is a compact Abstraction Network summarizing the structure (roles) of an ontology. It is composed of areas and child-of relationships connecting areas. An area is a group of concepts having the same set of role types. A concept can be in only one area, i.e., areas are disjoint. A concept that has no parent in its area, is called a *root* of the area. An area may have multiple roots. If a root concept of area B has a parent concept in area A, then there is a child-of relationship from area B to area A.

Fig. 1(a) is an excerpt of 12 Neoplasm concepts from NCIt. Concepts are represented as rounded-corner boxes and the arrows denote *IS-A* relationships. Concepts with the same set of role types are enclosed within a colored dashed rectangle. For example, both *Benign Neoplasm* and *Tumorlet* have the two role types *Disease Excludes Abnormal Cell* and *Disease Has Abnormal Cell*, they reside in the left green dashed rectangle. Fig. 1(b) shows the area taxonomy for Fig. 1(a). Each colored, dashed rectangle in Fig. 1(a) becomes an area with the same color in Fig. 1(b). An area is labeled by its role type set and the number of concepts it summarizes. Areas with the same number of role types have the same color. For example, there are two areas colored in green, since both have two role types. *Skin Neoplasm* is the root concept of the red area and its parent *Neoplasm by Site* is in the grey area. Hence, there is a child-of relationship from the red area to the grey area, denoted as a bold arrow in Fig. 1(b).

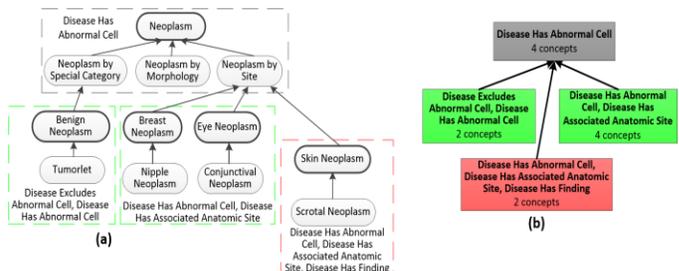


Fig. 1. (a) Excerpt of 12 concepts from the Neoplasm subhierarchy. (b) The area taxonomy for (a) with 4 areas.

## III. METHOD

We describe two methodologies, the unrestricted methodology and the refined restricted methodology. The ML training problem is viewed as a binary classification task: given a concept pair, we classify it into a positive category (there is an *IS-A* link) or a negative category (there is no *IS-A* link). We train a Convolutional Neural Network (CNN) model to solve this classification problem.

**Unrestricted Methodology:** To train a CNN model to yield high precision, we must carefully choose the training data for both categories. The source of training samples for the positive category is in the *IS-A* hierarchy of the ontology. The challenge is in the choice for the negative samples as we cannot use the full set of unconnected pairs. For the Neoplasm subhierarchy of NCIt with 9955 concepts, the size of this set is 99,075,537 ( $= 9955 * 9954 - 16533$ ). We subtract 16533 existing *IS-A* link pairs from the potential missing parent/child errors. Training pairs should not be chosen randomly. We need to choose pairs where there is a reasonable likelihood for an *IS-A* link, not pairs that obviously have no taxonomic relation. For example, a body part concept and a drug concept are not related by an *IS-A* relationship and would be a bad training pair. Negative training samples should be near misses, close to the “hyperplane of separation” in SVM terms.

To address these two problems of magnitude and recall in the ML training, we limit the negative samples for the unrestricted methodology to only “uncle - nephew” pairs for only near misses. That is, connections between a concept and a sibling of its parent. By only choosing “uncle - nephew” pairs, we guide the model to learn the underlying features used to distinguish *IS-A*-connected concept pairs and similarly positioned concept pairs that have a high potential to have secondary *IS-A* links but are not connected by *IS-A* links. There are total 37,147 such “uncle - nephew” pairs in the Neoplasm subhierarchy of NCIt.

The unrestricted model must consider all uncles of a concept (that are not connected to that concepts by an *IS-A* link). Due to the size of the Neoplasm subhierarchy, the number of uncles is large. For example, there are 24, 15 and 15 concepts with 10, 11 and 12 children, respectively and the maximum number of children of a concept is 60. Each grandchild of a concept with 15 children has 14 uncles. If the average number of children of each sibling is 5, then there are 70 concepts with 14 uncles each. Applying the proposed technique to select negative samples from the whole hierarchy results in a large number of “uncle - nephew” pairs. This is computationally expensive for training the model and leads to low accuracy by distracting the model from learning subtle features that distinguish between *IS-A* and “uncle - nephew” links within similar groups of concepts. In other words, not all “uncle - nephew” pairs are equally useful for training purposes. We need pairs that are similar to existing *IS-A*-connected pairs, yet are not themselves *IS-A*-connected.

Additionally, many machine learning models work best with balanced training sets. The number of positive and negative samples should be approximately equal. The number of positive training instances in our problem domain is given and fixed. The number of potential negative training samples is much larger. Thus a cogent way has to be used to select a number of negative training samples that is closer to the number of positive training samples.

**The Restricted Methodology:** To cope with these problems, we introduced the restricted approach. The restricted approach limits the number of negative samples by only choosing a

subset of a concept’s uncles, which are structurally similar to the investigated concept. To provide such a “closely related” subset we partition the Neoplasm subhierarchy into sets of concepts of similar structure. In doing this we are availing ourselves of a powerful mechanism that derives an *area taxonomy* from an ontology [3]. An area taxonomy is an Abstraction Network that clusters together groups of concepts according to their roles. All concepts in one area have exactly the same roles. Concepts in different areas differ in at least one role from each other.

In the area taxonomy each cluster of concepts constitutes an area. Due to the high average number of roles per concept in the Neoplasm subhierarchy, the number of areas is large, and the average size of an area is small. Note that AbNs are automatically derived from the ontology so the limitation process is automatic [22]. By selecting pairs only within areas, we narrow down 37,147 negative samples to 10,574 more closely related “uncle - nephew” pairs, of the same magnitude as the 16,533 positive samples.

Since all uncles in an area will have exactly the same roles as the current concept, they will be similar to it. Thus the recall of the CNN training model is expected to be higher than for the unrestricted model trained with all uncles of a concept, many of which are not similar to the current concept.

The following description is related to both methodologies. Overall, our methodology comprises following four steps:

#### A. Document Embedding

The CNN model requires its input in the format of fixed-length feature vectors. Thus, before sending concept pairs for training, we need to transform each concept into its corresponding vector representation with fixed length.

The Paragraph Vector (Doc2vec) framework introduced by [13] generates fixed-length feature vectors from variable-length pieces of text, as it was designed for text corpus processing. Thus, the problem to overcome in applying Doc2vec to ontologies is to find the vector representation of single concepts. However, an *IS-A* link is defined by a pair of concepts, thus a joint representation of pairs is needed that is also compatible with the input required by CNN.

To derive the vector representation of a single concept, we need text “descriptions” of the concept. We recast a concept into a *document* such that it preserves hierarchical and partial semantic information of the concept:

The *document* of a concept contains the concept ID, the name(s) of its ancestor(s), the name(s) of the concept itself, the name(s) of its child(ren) and the names of its grandchild(ren), if they exist. In this way, the document implicitly maintains the hierarchical relationships of the ontology.

For example, the document representation of the concept *Malignant Nipple Neoplasm* (Fig. 2) is “c5213: *Neoplasm* → *Neoplasm by site* → *Breast neoplasm* → *Malignant Breast Neoplasm* → *Nipple Neoplasm* → *Malignant Nipple Neoplasm* → *Female Malignant Nipple Neoplasm* → *Male Malignant*

*Nipple Neoplasm*  $\rightarrow$  *Nipple Carcinoma*.” Thus, the generated distributed vector representation maintains the most important hierarchical relationship semantics.

The document vectors are derived using the Distributed Memory version of Paragraph Vector (PV-DM) [13] via the Gensim [23] Doc2Vec implementation. Each vector has the dimensionality of 128. Pairs of concepts connected by an *IS-A* link are represented by the concatenation of the document vectors of the two concepts, with the child concept first.

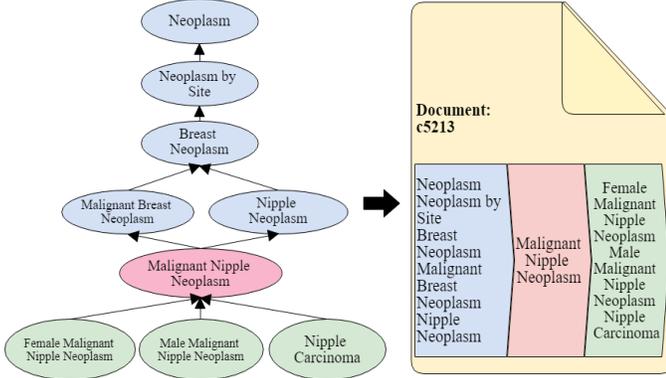


Fig. 2. Concept document derivation for *Malignant Nipple Neoplasm*

### B. Training and Testing Data

The positive samples are directly extracted from the hierarchy as all the concept pairs connected via an *IS-A* link. For example, (*Malignant Nipple Neoplasm*, *Nipple Neoplasm*) is a positive sample, because *Malignant Nipple Neoplasm* is a *Nipple Neoplasm* (Fig. 2). As mentioned above, there are 16,533 positive samples in the Neoplasm subhierarchy. We randomly picked 2,000 positive samples for testing. The remaining 14,533 (=16,533-2,000) samples are used in a ratio of 80% for training and 20% validation. The positive samples are treated equally for both models.

For the restricted model, there are 10,574 potential pairs where both concepts of each pair are from the same area. We randomly picked 2,000 for testing. As noted above, there are 37,147 “uncle – nephew” negative sample pairs in the hierarchy. However, for the unrestricted model, we use the same 2000 negative pairs, used for the restricted model. This is done to enable performance comparison between the two models. Similar to the way we handle the positive samples, the remaining 8,574 (=10,574-2,000) samples for the restricted model and the remaining 35,147 (=37,147-2,000) for the unrestricted model are divided to 80% vs. 20% ratio for training and validation, respectively.

In addition, we down-sampled the negative samples for the unrestricted model and up-sampled for the restricted model to 14,533 samples, in order to balance the number of samples for both categories, as customary in Machine Learning.

### C. CNN Model

We trained a CNN model with 4 convolution layers on top of vectors derived from the Neoplasm subhierarchy of NCI via

Doc2vec. The CNN model architecture is shown in Fig. 3. The input to the CNN model is two 128x1 dimension vectors and the output is a 2x1 dimension vector.

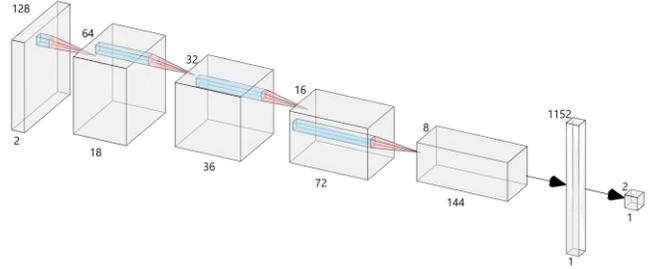


Fig. 3. CNN model architecture

Some structural details of this model are summarized as follows:

- There are four convolution layers, each followed by a max pooling layer. This choice was informed by previous research. The first convolution layer has 18 filters with kernel size =1. The filter number doubles with the increase of convolution layers. We use stride =1, meaning we slide the filters one number (position) at a time over the input. The pooling size is 2 for all max pooling layers.
- The Adam [24] optimization algorithm for stochastic gradient descent is used for training with the learning rate set to 0.001.
- The ReLU (Rectified Linear Unit) activation function is used in every convolution layer, because it has proven successful in recent research projects. This corresponds to a “rectifier function” from electrical engineering, blocking the negative half-wave and letting the positive half-wave pass through one-to-one.

### D. Test against Reviewed Data

Traditionally, machine learning models are tested with *k*-fold cross validation. Thus, all known data is partitioned into *k* folds, the model is trained with *k*-1 folds and tested with the remaining fold. This process is repeated *k* times, with resulting precision, recall and  $F_1$  values averaged. We have augmented this testing by human expert quality assurance results.

In a previous study [9], domain experts reviewed 190 concepts from the Neoplasm subhierarchy and reported 18 missing parent errors. This data was used as ground truth in this study to check the *sensitivity/recall* of our model’s performance.

## IV. RESULTS

We report our CNN model’s performance in the following three aspects:

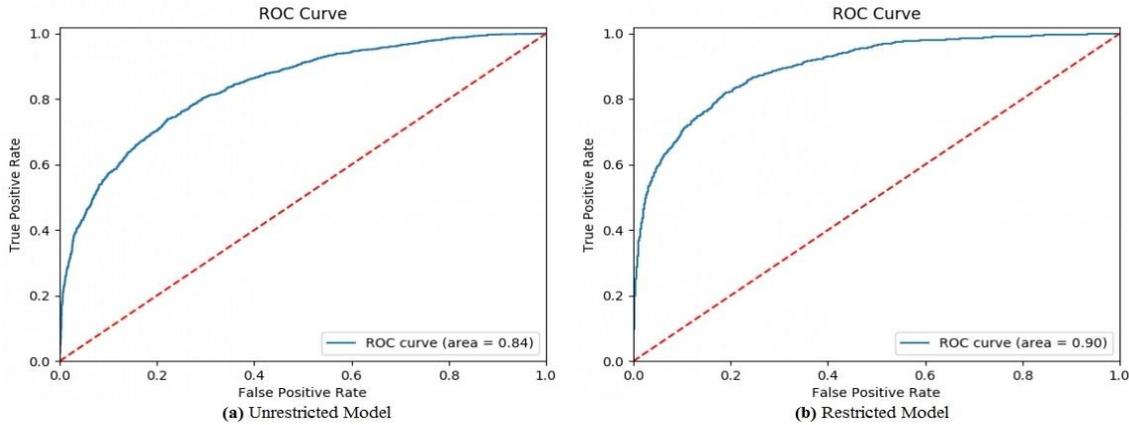


Fig. 4. ROC curve of the two models

### A. Testing recall and AUC

The testing recall is 0.75 and 0.81 for the unrestricted and restricted models, respectively. Fig. 4 (a) and (b) show the Receiver Operating Characteristic (ROC) curves of the testing performance for the unrestricted and restricted models, respectively. The AUC (area under the curve) scores, as the measure of test accuracy, are 0.84 and 0.90, respectively.

### B. Confirmed errors found by domain experts

The restricted model detected 10 out of the 18 errors that domain experts found [9], while the unrestricted model detected only five errors, all contained in the above 10 errors. Table I shows two missing parent examples confirmed by both models, and two examples confirmed only by the restricted model.

### C. Training time efficiency

Each model is trained with 2000 epochs, with batch size = 2000. We recorded the duration of the training. With the same computer hardware configuration, training the unrestricted and restricted models took 1116 and 1110 seconds, respectively.

TABLE I. MISSING PARENT ERRORS CONFIRMED BY THE TWO MODELS

Child	Model	Missing Parent
Reproductive Endocrine Neoplasm	Both	Endocrine Neoplasm
Basophilic Adenocarcinoma	Both	Anterior Pituitary Gland Neoplasm
Breast Tubular Adenoma	Restricted	Tubular Adenoma
Cutaneous Glomangioma	Restricted	Benign Skin Neoplasm

## V. DISCUSSION

To the best of our knowledge, this is the first published attempt to use Machine Learning (ML) for QA of ontologies. Such technique can prepare a subset of pairs of concepts as candidates for missing *IS-A* link omission errors, optimizing the use of scarce QA resources.

In this paper, we discussed two ML approaches. The unrestricted model utilizes all uncle concepts of the processed concept. The restricted model is further taking advantage of the area taxonomy of the ontology to utilize only uncles in the area

of the processed concept. Both ideas save processing time while improving the accuracy, by training unconnected pairs of concepts similar to the original *IS-A* links of the processed concept. The uncles are similarly positioned as siblings of the parent of the processed concept. The uncles within the area share the same roles as the processed concept.

Table II compares the performance of the two models. As can be expected, the restricted model utilizing training with similar concepts achieves higher performance. We evaluated the results of the two ML models based on a list of errors that domain experts found in a previous study [9]. Such an evaluation is usually not available for ML studies. An interesting observation is that the set of five errors confirmed by the unrestricted model is a subset of the 10 errors found by the restricted model. The reason for this may be that the two models use the same negative test pairs where the uncles are from the same area as the nephew. This choice was made in order to be able to compare performance of the two models, but tends to unnecessarily limit the unrestricted model to find error pairs of similar concepts. In the future, we will perform experiments where the unrestricted and restricted models will use disjoint negative training data in an effort to optimize the results of each model rather than to compare them.

TABLE II. TWO MODELS PERFORMANCE COMPARISON

	Unrestricted Model	Restricted Model	Difference
<b>Confirmed Errors</b>	5	10	5
<b>Corresponding Recall (out of 18)</b>	0.28	0.56	0.28
<b>Testing Recall</b>	0.75	0.81	0.06
<b>Testing AUC</b>	0.84	0.90	0.06
<b>Training Time (sec)</b>	1116	1110	6

From Table II we can see that the restricted model performs 6% better than the unrestricted model in recall. The area under the curve in Fig. 4(b) is 6% larger than in Fig. 4(a), reflecting a better classification of the restricted model than the unrestricted model. The document vectors used in this study were derived using the Distributed Memory version of Paragraph Vector (PV-DM) that works well for most tasks, as stated in the original paper [13]. However, it is also recommended in the paper to combine Paragraph Vector with Distributed Bag of

Words (PV-DBOW) to obtain consistency. The more accurate the vector representations of concepts are, the better recall should be expected. This is left for future work.

The recall obtained is not high enough for reliable QA for missing *IS-A* links. For example, out of a random subset of 20 suggested errors, a domain expert (GE) confirmed only one error pair, *Hair Follicle Neoplasm* missing the parent *Dermal Neoplasm*, since hair follicles reside in the dermal layer of the skin. In the NCI, *Dermal Neoplasm* has 13 children, representing a mix, based on cell origin as well as malignancy status. Currently, *Hair Follicle Neoplasm* has *Skin Appendage Neoplasm* as a parent. *Dermal Neoplasm* and *Skin Appendage Neoplasm* are siblings. However, *Hair Follicle Neoplasm* has only a very indirect relationship to the dermis through the *Disease Has Primary Anatomic Site* role with *Hair Follicle* as the target value and the *Anatomic Structure Is Physical Part Of* role of *Hair Follicle* with the target value of *Dermis*. Adding *Dermal Neoplasm* as a parent, or even possibly replacing *Skin Appendage Neoplasm* with it, might be better.

A higher recall will imply fewer suggested errors with a higher percentage of confirmed errors by domain experts. In future research, we will further explore the properties of the two models as well as properties of ML processing in an effort to best fine-tune and utilize both models, to increase the recall.

## VI. LIMITATIONS

We presented a supervised training technique, therefore “annotated corpora” are required for its applicability beyond finding missing *IS-A* relationships in the NCI neoplasm taxonomy. The results are also limited, because the models are not trained on more general data for more general problems.

## VII. CONCLUSION

We explored whether ML methods can be applied to the task of ontological QA, in particular whether ML can help in detecting missing *IS-A* links in an ontology. Two models are presented. Our application of ML to the Neoplasm subhierarchy of NCI demonstrated that the restricted model performs better than the unrestricted one. However, the performance of the restricted model is not yet sufficient for QA. In future research, we will explore improvements in ML processing and more accurate restrictions for concepts in the training stage to improve performance of QA of ontologies.

## ACKNOWLEDGMENT

Research reported in this publication was partially supported by the National Cancer Institute of the National Institutes of Health under Award Number R01CA190779. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## REFERENCES

[1] S. de Coronado, M. W. Haber, N. Sioutos, M. S. Tuttle, and L. W. Wright, "NCI Thesaurus: using science-based terminology to integrate

cancer research results," *Stud Health Technol Inform*, vol. 107, no. Pt 1, pp. 33-7, 2004.

[2] G. Elhanan, Y. Perl, and J. Geller, "A survey of SNOMED CT direct users, 2010: impressions and preferences regarding content and quality," *J Am Med Inform Assoc*, vol. 18 Suppl 1, pp. i36-44, 2011.

[3] H. Min, Y. Perl, Y. Chen, M. Halper, J. Geller, and Y. Wang, "Auditing as Part of the Terminology Design Life Cycle," *J Am Med Inform Assoc*, vol. 13, no. 6, pp. 676-690, Nov-Dec 2006.

[4] X. Zhu, J. W. Fan, D. M. Baorto, C. Weng, and J. J. Cimino, "A review of auditing methods applied to the content of controlled biomedical terminologies," *J Biomed Inform*, vol. 42, no. 3, pp. 413-25, Jun 2009.

[5] J. Geller, Y. Perl, M. Halper, and R. Cornet, "Special issue on auditing of terminologies," *J Biomed Inform*, vol. 42, no. 3, pp. 407-11, 2009.

[6] M. Arguello Casteleiro *et al.*, "A case study on sepsis using PubMed and Deep Learning for ontology learning," *Informatics for Health: Connected Citizen-Led Wellness and Population Health*, vol. 235, pp. 516-520, 2017.

[7] J. A. Minarro-Giménez, O. Marin-Alonso, and M. Samwald, "Exploring the application of deep learning techniques on medical text corpora," *Stud Health Technol Inform*, vol. 205, pp. 584-588, 2014.

[8] İ. Pembeci, "Using Word Embeddings for Ontology Enrichment," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 4, no. 3, pp. 49-56, 2016.

[9] L. Zheng, H. Min, Y. Chen, J. Xu, J. Geller, and Y. Perl, "Auditing National Cancer Institute thesaurus neoplasm concepts in groups of high error concentration," *Appl Ontol*, vol. 12, no. 2, pp. 113-130, 2017.

[10] M. Halper, H. Gu, Y. Perl, and C. Ochs, "Abstraction networks for terminologies: Supporting management of "big knowledge", " *Artif Intell Med*, vol. 64, no. 1, pp. 1-16, May 2015.

[11] Y. Wang, M. Halper, D. Wei, Y. Perl, and J. Geller, "Abstraction of complex concepts with a refined partial-area taxonomy of SNOMED," *J Biomed Inform*, vol. 45, no. 1, pp. 15-29, Feb 2012.

[12] Y. Wang *et al.*, "Auditing Complex Concepts of SNOMED using a Refined Hierarchical Abstraction Network," *Journal of Biomedical Informatics*, vol. 45, no. 1, pp. 1-14, 09/01 2012.

[13] Q. V. Le and T. Mikolov, "Distributed Representations of Sentences and Documents," *CoRR*, vol. abs/1405.4053, / 2014.

[14] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv:1301.3781*, 2013.

[15] Y. Shen, X. He, J. Gao, L. Deng, and G. Mesnil, "Learning semantic representations using convolutional neural networks for web search," in *Proc. of the 23rd Int. Conf. World Wide Web*, 2014, pp. 373-374: ACM.

[16] W.-t. Yih, X. He, and C. Meek, "Semantic parsing for single-relation question answering," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2014, vol. 2, pp. 643-648.

[17] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," *arXiv preprint arXiv:1404.2188*, 2014.

[18] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.

[19] Y. Lin, S. Shen, Z. Liu, H. Luan, and M. Sun, "Neural relation extraction with selective attention over instances," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2016, vol. 1, pp. 2124-2133.

[20] Y. Wang, M. Halper, H. Min, Y. Perl, Y. Chen, and K. A. Spackman, "Structural methodologies for auditing SNOMED," *Journal of biomedical informatics*, vol. 40, no. 5, pp. 561-581, 2007.

[21] H. Min, L. Zheng, Y. Perl, M. Halper, S. de Coronado, and C. Ochs, "Relating Complexity and Error Rates of Ontology Concepts," *Methods of Information in Medicine*, vol. 56, no. 03, pp. 200-208, 2017.

[22] C. Ochs, J. Geller, Y. Perl, and M. A. Musen, "A unified software framework for deriving, visualizing, and exploring abstraction networks for ontologies," *JBI*, vol. 62, pp. 90-105, 2016/08/01/ 2016.

[23] R. Rehurek and P. Sojka, "Software framework for topic modelling with large corpora," in *In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 2010: Citeseer.

[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.