

Detecting Anatomical and Functional Connectivity Relations in Biomedical Literature via Language Representation Models

Ibrahim Burak Ozyurt
FDI Lab Dept. of Neuroscience
UCSD La Jolla, USA
iozyurt@ucsd.edu

Joseph Menke
Scicrunch.com
San Diego, USA
joe@scicrunch.com

Anita Bandrowski
FDI Lab Dept. of Neuroscience
UCSD La Jolla, USA
abandrowski@ucsd.edu

Maryann E. Martone
FDI Lab Dept. of Neuroscience
UCSD La Jolla, USA
maryann@ncmir.ucsd.edu

Abstract

Understanding of nerve-organ interactions is crucial to facilitate the development of effective bioelectronic treatments. Towards the end of developing a systematized and computable wiring diagram of the autonomic nervous system (ANS), we introduce a curated ANS connectivity corpus together with several neural language representation model based connectivity relation extraction systems. We also show that active learning guided curation for labeled corpus expansion significantly outperforms randomly selecting connectivity relation candidates minimizing curation effort. Our final relation extraction system achieves $F_1 = 72.8\%$ on anatomical connectivity and $F_1 = 74.6\%$ on functional connectivity relation extraction.

1 Introduction

The NIH Common Fund’s Stimulating Peripheral Activity to Relieve Conditions (SPARC) program aims to transform our understanding of nerve-organ interactions to help spur the development of effective bioelectronic treatments. Bioelectronic medicine represents the convergence of molecular medicine, neuroscience, engineering and computing to develop devices to diagnose and treat diseases (Olofsson and Tracey, 2017). One of the projects within this large consortium is to create a systematized and computable wiring diagram of the autonomic nervous system, a part of the “wiring system” that travels throughout the body transmitting messages between the peripheral organs and the brain or spinal cord. While diagrams of nerves are currently available in medical texts (Strandring and Gray, 2008), the SPARC program seeks to map these connections at higher levels of detail and with greater accuracy. Additionally, the diagrams in these medical texts are not generally queryable, nor

are they sufficiently detailed to include the granular paths that these nerves travel. Such information would be needed, for example, to understand where reliable access points to a particular nerve might be so that stimulation only affects the most relevant nerve or to understand the mechanisms behind stimulation applied at particular locations. Many scientific studies contain information about individual nerves and at times the paths they traverse, but to our knowledge, no systematic approach has been attempted to bring these large quantities of information together into a computationally accessible format.

The SPARC project is building a cross-species connectivity knowledge base that contains detailed information about individual nerves, their pathways, cells of origin and synaptic targets. To date, this knowledge base has been populated through the development of detailed models of circuitry by experts funded through the SPARC project using the ApiNATOMY platform (Kokash and de Bono, 2021). ApiNATOMY provides a modeling language for representing the complexity of functional and anatomical circuitry in a standardized form. The circuitry contained in these models represent expert knowledge derived from the synthesis of the expert’s own work and the synthesis of, in some cases, hundreds of scientific publications. However, to ensure that information in the SPARC knowledge base is comprehensive and up to date, i.e., it represents the current state of knowledge about autonomic nervous system (ANS) connectivity, we sought to augment the expert-based model approach with experimental information derived from the primary scientific literature. As there are thousands of papers and additional sources like text books, we utilized natural language processing to identify sentences that contained information on neuronal connectivity in the ANS.

The task was approached by first gathering the relevant scientific literature by matching bodily structures at a variety of anatomical levels (i.e. gasserian ganglion, vagus nerve, brainstem, etc.) from a constructed set of vocabulary at sentence level. Then, annotators classified each structure to structure relationship using only the information provided within the sentence based on the connectivity types defined in our annotation guideline. This structured data were then used to train our connectivity relations models. Data from two curators was used to assess the inter-curator agreement to determine if the annotation guidelines are sufficient to “teach” the task to humans. We assessed connectivity statements into several types including, anatomical connectivity, functional connectivity, structural connectivity, topological connectivity and general connectivity as well as no connectivity. The general connectivity and no connectivity categories can be thought of as statements that are too vague to be of much direct use for our use case. The most important statements are anatomical connectivity, elucidating which parts are connected physically and functional connectivity, elucidating which parts are connected functionally. A definition and an example for each connectivity type used for annotation is shown in Table 1. Of course with single sentences, it is difficult to define a direct functional relationship, which typically rests on the latency with which a signal is detected between two elements (Bennett, 2001). However, statements about latency are very rare in the subset of the peripheral nervous system literature, whereas somewhat more general statements about functional relationships that, for example, describe damage to one area and altered functioning in another, are more abundant. We hypothesize that when such statements are reasonably abundant, a detection classifier will be easier to train.

In relation extraction, long-range relations are usually handled using dependency parse tree information. In traditional feature-based models, paths in the dependency parse tree between entities are used as features (Kambhatla, 2004) which suffered from the sparsity of the feature patterns. More recently, neural models are increasingly employed for relation extraction instead of feature engineering using vectorized word embeddings. The dependency information is represented as computation graphs along the parse tree (Zhang et al., 2018). Sequence models, on the other hand, work at the

surface level and represent long distance relationships via either convolutional or recurrent neural networks and an attention mechanism (Zhang et al., 2017).

In biomedical domain, relation extraction work is traditionally focused on protein-protein, gene-disease or protein-chemical interactions. Several labeled datasets, such as GAD (Bravo et al., 2015) (a gene-disease relation dataset) and CHEMPROT (Krallinger et al., 2017) (a protein-chemical multi-relation dataset) are publicly available. Neural sequence models have also been applied to protein-chemical relation extraction task (Lim and Kang, 2018).

Recently, sentence level transformer based language representation models such as BERT (Devlin et al., 2019) have shown superior downstream performance on many NLP tasks. A biomedical domain adapted version of BERT called BioBERT (Lee et al., 2019) has been shown state of the art performance on several biomedical relation extraction tasks.

While most of the transformer based language representation models are pretrained on sentences where a predefined percentage of the tokens are masked and the model learns to predict the masked tokens, a recently introduced language representation model, ELECTRA (Clark et al., 2020) learns to discriminate if a token in the original input is replaced by a language generator model or not. The generator model is a BERT like generative model that is co-trained with the discriminative model.

While there are efforts to extract brain connectivity information from neuroscience literature (Richardet et al., 2015), their focus is in the cognitive parts of the brain instead of ANS. In this paper, we introduce a labeled ANS connectivity corpus, together with four biomedical domain adapted ELECTRA models, that we have used to develop an anatomical and functional connectivity relation extraction system that outperforms BioBERT.

2 Methods

2.1 Vocabulary

In order to better structure information from papers, anatomical structure labels were drawn from a set of relevant ontologies, also approved for use by the SPARC project. These ontology terms include primarily FMA (RRID:SCR_003379), UBERON (RRID:SCR_010668), and NIFSTD (RRID:SCR_005414) terms, and they are listed

Relation	Definition	Example
functional	a relationship was determined to exist between two structures using physiological techniques	The HB reflex is a reflex initiated by lung inflation, which excited the myelinated fibers of vagus nerve , pulmonary stretch receptors [11,19].
anatomical	a physical synaptic relationship was observed between two structures using anatomical techniques such as tract tracing	Only the most prominent nervous connections, such as the penis nerve cord (pnc, Fig. 8a), connecting the ventral ganglion to the penis ganglion can be detected.
structural	a relationship that reflected continuity between segments of nerves	The term vocal fold paralysis (VFP) refers to the reduced or absent function of the vagus nerve or its distal branch, the recurrent laryngeal nerve (RLN) [1-3].
topological	a relationship that reflected the course of a nerve	Oculomotor nerve (III) exited from the middle tectum nearby ventro-medial midbrain and was observed on 6-day-old fish.
general	a statement that contained general information about connectivity but did not specify the technique used or otherwise failed to elucidate the exact type of connectivity discussed	Moreover, an interoceptive circuit connecting the gut to the nucleus tractus solitarius (NTS) via the vagus nerve has been demonstrated to convey the state of the gut to the limbic system (Figure 9; Maniscalco and Rinaman, 2018).

Table 1: Connectivity relation types

on the SPARC anatomy working group web pages, which include term lists. In order to provide a more targeted set of sentences for training, we selected a set of terms that was specifically associated with the ANS. These terms included sympathetic and parasympathetic nerves and ganglia from the FMA and UBERON. Terms were selected by the SPARC Anatomical Working Group, a group of anatomical experts who provide expertise to the SPARC knowledge engineers.

2.2 Corpus Generation

The sentences of interest for connectivity relation extraction were detected by longest phrase match from the target vocabulary of anatomical terms. We have used four million full-length PMC open access papers downloaded on November 2020 to search for sentences of interest. All the sentences mentioning at least two distinct anatomical structures from our vocabulary are selected. To focus our curation effort to a manageable portion of the vocabulary, a smaller vocabulary consisting of only ANS nerve and ganglion terms were selected to further filter the candidate sentence set where only sentences having at least one structure from the focused vocabulary set is selected. Since the resulting candidate set was still too large for curation, up to three examples from each unique focused

vocabulary term encountered is randomly sampled to create our base corpus of 808 sentences to be curated.

2.2.1 Annotation of the Corpus and Inter-annotator Agreement

Three curators/domain experts were involved in the connectivity corpus labeling process. Our main curator (J.M.) is a full-time curator with several years of experience with biomedical named entity recognition, relation extraction and text classification curation tasks. Dr. A.B. is a neurophysiologist by training with expertise in microcircuitry. Dr. M.M. is a trained anatomist with expertise in microcircuitry and a professor of neuroscience. All curators annotated training sets using the relation annotator tool developed in-house. The tool allows the curators both to edit entities (if automatically detected anatomical structure boundaries are not correct) and to label automatically generated binary relation combinations arising from the two or more anatomical structures detected in the curated sentence.

To start, A.B. and J.M. completed 30 sentences together to gauge the difficulty of the task and train J.M. on the differences between anatomical and functional connectivity. Then, M.M. and J.M. independently annotated 102 relation labels. In this first iteration, connectivity between structures

could only be classified as anatomical connectivity, functional connectivity, or no relation. The initial inter-annotator agreement was 66.7%; Cohen’s kappa was 0.25. Even when simply comparing binary connectivity vs. no relation, there was only an inter-annotator agreement of 72%; Cohen’s kappa was 0.34. After discussing disagreements, additional connectivity types were added to the relation annotator tool. After expansion, connectivity could be classified as structural, topological, or general in addition to the previous versions labels: anatomical, functional, and no relation. This was done to make each connectivity type more explicit with less potential overlap, especially between our main connectivity types of interest (anatomical vs. functional).

In our second iteration, M.M. and J.M. independently annotated another 170 relation labels across 100 sentences. This time, the inter-annotator agreement was 73.5%; Cohen’s kappa was 0.25. Annotation differences though were primarily found to be between our general connectivity and no relation tags. If we consider general connectivity to be the same as no relation (collapsing them together), then our inter-annotator agreement jumps to 91.2%; likewise, Cohen’s kappa also increases to 0.55. Because our primary disagreements were between two tags of less interest (general connectivity vs. no relation), we believe our inter-annotator agreement is acceptable for this high difficulty task.

2.3 Models

2.3.1 ELECTRA based language representation models for Biomedical Domain

Domain specific language representation models result in performance improvements on downstream NLP tasks as demonstrated by BioBERT (Lee et al., 2019). Similarly, we have pretrained four ELECTRA (Clark et al., 2020) based models on biomedical corpus.

For pretraining corpus we have used both PubMed abstracts and PubMed Central (PMC) open access full-length papers. 21.2 million PubMed abstracts from the January 2021 baseline distribution are used to build our main pretraining corpus. Sentences extracted from the paper title and abstract text resulted in a corpus of 3.6 billion words. For the PMC open access papers, sentences extracted from all sections except the references section of the full-length papers are used to build a

12.3 billion words corpus. A domain specific word piece vocabulary is generated using SentencePiece byte-pair-encoding (BPE) model (Sennrich et al., 2016) from PubMed abstract texts. The models are pretrained for one million steps on the PubMed abstracts corpus followed by 200,000 steps training on the PMC open access papers corpus.

During training, ELECTRA uses a small transformers based encoder model using masked language objective like in BERT to generate possible replacements for the larger discriminative model which is also based on transformers architecture. Both models are trained jointly. During fine-tuning, only the discriminative model parameters are used. The discriminative model has essentially the same architecture as BERT but trained in a discriminative manner using a different objective. We have trained three different model sizes; a base model with embedding and hidden size of 768, 12 attention heads and 12 transformer layers; a mid sized model with embedding size of 384, hidden size of 512, 8 attention heads and 12 transformer layers; a mid sizes tall model having same parameters as the mid sized model but with 24 transformer layers. We have also trained another mid sized model with the combined PubMed abstract and PMC open access full paper corpus instead of the two corpus cascaded training approach used for the other three Bio-ELECTRA models. For all models, the maximum allowed input sequence length was set to 512. For all models besides the mid-tall model, the batch size was set to 256. The mid-tall model had a batch size of 128 because of the memory limitations of a single tensor processing unit (TPU). The model architectures, sizes and training times are summarized in Table 2. All the models are trained on a single 8 core version 3 TPU with 128 GB RAM.

While, we have used our Bio-ELECTRA models for connectivity relation extraction only, the models like BioBERT are applicable to many downstream biomedical NLP tasks.

3 Experiments

We conducted our experiments in two phases. In the first phase, all the binary connectivity relation candidates in the 805 sentences extracted from the open access subset of PubMed Central is annotated by a curator. The curated base set is then randomly split into 80/20% train/test set. Afterwards, ten randomly initialized models are trained. The reported results are average of 10 runs together with

Model	Params	Architecture	Steps	Train Time/Hardware
Mid	50M	hidden:512, layers:12	1.2M	6.5d on 8 TPUv3s
Base	110M	hidden:768, layers:12	1.2M	12.5d on 8 TPUv3s
Mid-tall	88M	hidden:512, layers:24, batch:128	1M	5.5d on 8 TPUv3s
Mid Combined	50M	hidden:512, layers:12	1.2M	6.5d on 8 TPUv3s

Table 2: ELECTRA Models for Biomedical Domain

standard deviation. In the second phase, the base set is enhanced via active learning.

As our baseline model, we have used a graph convolution over dependency parse tree neural model (Zhang et al., 2018) where the dependency graph structure is represented by an adjacency matrix over which convolution operations are performed. The model uses word embedding vectors for input encoding and stacked layers of graph convolution network (GCN) layers to encode relations. The input encodings can be further contextualized via a bi-directional long-short-term memory (LSTM) layer, which we have used in our experiments. For word embeddings we have used 300 dimensional Common Crawl (840B tokens) trained GloVe (Pennington et al., 2014) vectors. The dependency parse trees for the input sentences were generated via Stanford CoreNLP (Manning et al., 2014) package.

All the other models are fine-tuned from pre-trained transformers based language representation models. We have downloaded Bio-ELECTRA++ from Zenodo¹. Besides our four biomedical corpus pretrained ELECTRA models, we have used BioBERT (Lee et al., 2019) version 1.1 and ELECTRA Base models. The binary anatomical structure entities are masked in candidate sentences as in (Zhang et al., 2018; Lee et al., 2019). Besides that, no further preprocessing is done. All the models are trained for three epochs, using the the default learning rate and maximum allowed batch size for our 8GB Nvidia RTX 2070 GPU.

The test performance of models tested are summarized in Table 3. Even after the benefit of dependency parses, contextualized graph convolution networks were at the bottom of the performance rank tying with the smallest language representation model. Two Bio-ELECTRA models, namely Bio-ELECTRA Base and Bio-ELECTRA Mid outperformed BioBERT. Given that the Bio-ELECTRA Mid has less than half the parameters of Bio-BERT,

its performance is especially impressive. We chose the best performing Bio-ELECTRA Base model for the second stage.

3.1 Extending Curation Set via Active Learning

Since labeled data set generation is costly and time consuming, we have tried to leverage active learning to minimize curation effort while trying to maximize prediction performance. To this end, 250 randomly selected candidate sentences from the nerve-ganglia PMC data set, are interactively curated by our curator in ten iterations. Each iteration has consisted of 25 candidate sentences selected by the binary relation extraction classifier trained on all the the curated sentences from the previous iterations plus the base training set. In the first iteration, the classifier is trained on the base training set only. For the control set, we have randomly selected 250 candidate sentences from the nerve-ganglia PMC data set, which are annotated separately by our curator. We have used uncertainty sampling as our oracle query strategy where the 25 unlabeled sentences that are closest to the decision boundary (probability estimate of 0.5) are selected for curation at each iteration. After each iteration, the extended training set is used to train ten randomly initialized models which are tested on the testing set. The precision and F_1 performance scores over the active learning set is shown in Figure 1.

The testing performance of active learning based vs random selection based training set expansion is shown in Table 4. Active learning strategy was significantly better than random selection based on two-tailed t test.

3.2 Effect of Hyperparameter Optimization

The additional 500 curated sentences (250 from active learning, 250 from random control set) are combined with the base training set. To maximize relation extraction performance, we used hyperparameter tuning on the 80%/20% training/dev set split of the combined training set. Using hyper-

¹<https://doi.org/10.5281/zenodo.3971235>

Model	Parameters	Precision	Recall	F_1
Contextualized-GCN		71.05 (4.36)	54.23 (4.20)	61.36 (3.01)
ELECTRA Base	110M	69.35 (4.23)	70.85 (5.43)	70.03 (4.39)
BioBERT	110M	67.82 (4.71)	72.34 (2.18)	69.89 (2.40)
Bio-ELECTRA++	11M	54.41 (2.11)	70.32 (3.38)	61.26 (1.33)
Bio-ELECTRA Mid	50M	69.16 (3.53)	73.83 (2.24)	71.36 (2.16)
Bio-ELECTRA Base	110M	69.93 (2.91)	74.26 (3.55)	71.99 (2.76)
Bio-ELECTRA Mid Combined	50M	67.66 (2.38)	74.36 (5.80)	70.70 (2.78)
Bio-ELECTRA Mid-tall	88M	63.89 (4.51)	65.96 (3.81)	64.78 (2.98)

Table 3: Binary connectivity/no-connectivity relation extraction on base set

Data Set	Precision	Recall	F_1
Random	70.29 (1.69)	74.04 (3.27)	72.06 (1.68)
Active learning	75.88 (2.70)	75.11 (2.39)	75.47 (2.30)

Table 4: Test performance effect for active learning vs random selection based labeled set expansion

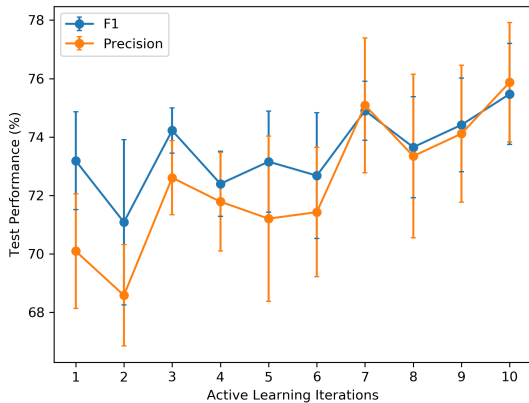


Figure 1: Average test performance over active learning iterations

opt (Bergstra et al., 2013) Python package, we searched for the optimum F_1 value for the following hyperparameters; the learning rate among the values $1e-5$, $5e-5$, $1e-4$ and $5e-4$, number of epochs among the values 3, 5, 10. We have used the maximum possible batch size of 16 for our 8GB RAM GPU. The best performing hyperparameter combination was then used to train ten randomly initialized Bio-ELECTRA Base based connectivity relation extraction classifiers. The results together with ten runs using default learning parameters are shown in Table 5. Hyperparameter optimization performance was significantly better than default learning parameter performance as determined by two-tailed t test ($p = 0.014$)

To detect anatomical and functional connectivity relations among candidate structure binary relation

sentences, we have introduced a three class classifier based on the same Bio-ELECTRA Base language representation model as the connectivity/no-connectivity classifier. Ten randomly initialized classifiers are trained using optimized hyperparameters. The test performance is shown in Table 6.

4 Discussion

Connectivity relations constituted only about 12% of the connectivity relationship candidates in our corpus. Taking this into account, the anatomical and functional connectivity detection performance of our final classifier is good enough to be used for ANS connectivity knowledge base construction with drastically reduced domain expert curation.

When looking at our model’s performance, we considered errors at the level of individual connectivity relations labels, meaning we could (and did) have some sentences with multiple errors. We defined errors as cases where relation labels tagged by the model and the annotator did not agree. We performed our error analysis in two phases. In phase 1, our analysis was performed using only binary connectivity data (i.e. did annotators mark a relationship using any type of connectivity or as no relation) from 40 connectivity errors: 19 false positives and 21 false negatives. False positives were defined as instances when the model predicted connectivity when there was actually no relationship (as defined by the annotator). Inversely, false negatives were cases when the model predicted structures to have no relationship when there was actually some type of connectivity. In phase 2, our

Model	Precision	Recall	F_1
Bio-ELECTRA Base (default)	76.97 (2.72)	74.68 (2.32)	75.77 (1.99)
Bio-ELECTRA Base (opt)	77.32 (2.39)	77.98 (1.65)	77.62 (1.33)

Table 5: Connectivity/No-Relation test performance on the extended training set

Relation	Precision	Recall	F_1
Anatomical connectivity	68.93 (2.94)	77.12 (1.37)	72.77 (1.99)
Functional connectivity	82.79 (2.39)	68.00 (2.80)	74.61 (2.35)

Table 6: Anatomical/functional connectivity test performance on the extended training set

analysis was performed on anatomical connectivity errors (18 false positives; 15 false negatives) and functional connectivity errors (5 false positives; 11 false negatives). With both phases, we noticed patterns emerging among the errors, although in most cases, these errors were present across all connectivity types. In other words, there was very little difference between the errors seen in phase 1 vs. the errors seen in phase 2.

The first identified error pattern was mislabeled data due to human error. We noticed 4 instances where the data was mislabeled. The second type of error occurred because a solid line of demarcation between connectivity types was difficult to establish due to ambiguities in our curation guidelines and the overall difficulty of the task. When we began annotating, connectivity between structures could only be defined as anatomical, functional or having no relation. After discussing the differences in our annotations though, it became apparent we needed to add additional connectivity types to clarify the lines of demarcation between each. As a result, we added structural, topological, and general connectivities, and while we did see improved classifier performance after adding these, it appears we weren't entirely successful in our attempts to explicate our connectivity types. In the example,

"The central amygdaloid nucleus (CeA) and the **bed nucleus of the stria terminalis** (BNST), which is considered to be a component of the "**extended amygdala**", establish important connections with the hypothalamus and other brain areas controlling visceral and sensory information."

BNST and the extended amygdala were incorrectly identified as having anatomical connectivity. This "part of the whole" pattern was seen multiple times in our errors (binary, anatomical and functional),

primarily as a false positive. In future works, it might benefit curation efforts to add additional connectivity type(s), e.g. fractional connectivity for this "part of the whole" pattern, in order to further elucidate lines of demarcation. Additionally, because connections between structures are not always obvious, even to human curators (i.e. if A is connected to B and B is connected to C, is A always connected to C?), the lines of demarcation separating connectivity types may always remain somewhat hazy.

We were able to identify a few additional patterns by comparing the syntax and vocabulary of sentences with errors to that of sentences without errors. In general though, sentences with errors tended to have far more complex sentence structure than sentences without. More specifically, error prone sentences generally contained far more prepositional phrases and compound subject and verb phrases. For example, we saw multiple errors within the following sentence:

"Chemoreceptors in the carotid body or aortic body in the walls of the internal carotid artery or the aorta sense the level of oxygen or carbon dioxide in the blood and convey these signals via the glossopharyngeal and vagus nerves to the nucleus of the tractus solitarius."

Just from a cursory glance, it becomes obvious that this sentence is complicated; it contains multiple subject and verb phrases clouded by prepositional phrases. Unfortunately, the convoluted nature of the sentence hurts readability for both humans and machines. Because humans also tend to have issues understanding these highly complex sentences, we feel the best solution is for authors to limit the complexity of their sentences to reasonable levels when possible. If a sentence is too complex for a human to understand, it will most likely be too complex

for a computer. Additionally, we noticed persistent issues when subjects were not explicit. Unresolved pronouns (e.g. pronouns whose antecedents are unknown) and ambiguous body structures (i.e. fibers) tended to cause errors wherein the model would correctly identify that the sentence contained connectivity but would incorrectly identify which structures are connected. With regards to verb usage, our model seemed to perform better when the connectivity between structures was described in active voice rather than passive. One potential explanation is that sentences using active voice tend to be more clear and simple than sentences using passive voice. Lastly, our model seemed to perform worse the further apart the two connecting structures were within the sentence.

5 Conclusions

In this paper, we introduced a labeled corpus for ANS connectivity relations which is further expanded via active learning. The labeled ANS connectivity relation corpus is used to develop relation extraction systems mostly based on language representation neural models. We have introduced four biomedical domain pretrained ELECTRA (Clark et al., 2020) based discriminative language representation models, two of which have outperformed BioBERT (Lee et al., 2019) on the ANS connectivity relation extraction task. Using active learning guided curation, the labeled corpus is expanded minimizing the curation effort while significantly improving ANS connectivity relation extraction performance.

Based on the observed benefits of the active learning, we are planning to use our Bio-ELECTRA based relation extraction system in a web based tool for ANS connectivity knowledge base construction with active learning based continuous learning ability.

Software and Data Availability

All pretrained Bio-ELECTRA models are available on Zenodo (<https://doi.org/10.5281/zenodo.4699034>). The labeled connectivity corpus and codebase including the connectivity relation annotation tool are available on Github (<https://github.com/SciCrunch/connectivity-re>).

Acknowledgements

This work was conducted under the auspices of the Stimulating Peripheral Activity to Relieve Conditions (SPARC) program (RRID:SCR_017041) supported by the NIH Common fund award numbers 1OT3OD025349 and K-CORE 1OT2OD030541. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We would also like to thank Google TensorFlow Research Cloud (TFRC) program for providing us with free TPUs which allowed us to pretrain our ELECTRA models.

Conflicts of interest

AB, MM, and IBO are co-founders of SciCrunch Inc, a company devoted to improving scholarly communication.

References

- M. R. Bennett. 2001. *History of the Synapse*. CRC Press.
- J. Bergstra, D. Yamins, and D. D. Cox. 2013. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28, ICML'13*, page I-115-I-123. JMLR.org.
- Àlex Bravo, Janet Piñero, Núria Queralt-Rosinach, Michael Rautschka, and Laura I Furlong. 2015. Extraction of relations between genes and diseases from text and large-scale data analysis: implications for translational research. *BMC bioinformatics*, 16(1):55.
- Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. *Electra: Pre-training text encoders as discriminators rather than generators*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. *BERT: Pre-training of deep bidirectional transformers for language understanding*. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171-4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Nanda Kambhatla. 2004. *Combining lexical, syntactic, and semantic features with maximum entropy models for information extraction*. In *Proceedings of the ACL Interactive Poster and Demonstration Sessions*,

- pages 178–181, Barcelona, Spain. Association for Computational Linguistics.
- N. Kokash and B. de Bono. 2021. [Knowledge representation for multi-scale physiology route modeling](#). *Frontiers in Neuroinformatics*, 15(560050).
- Martin Krallinger et al. 2017. Overview of the BioCreative VI chemical-protein interaction track. In *Proceedings of the BioCreative VI Workshop*, pages 141–146, Bethesda, MD.
- Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2019. [BioBERT: a pre-trained biomedical language representation model for biomedical text mining](#). *Bioinformatics*, 36(4):1234–1240.
- Sangrak Lim and Jaewoo Kang. 2018. [Chemical-gene relation extraction using recursive neural network](#). *Database*, 2018. Bay060.
- Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. [The Stanford CoreNLP natural language processing toolkit](#). In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60.
- P. S. Olofsson and K. J. Tracey. 2017. [Bioelectronic medicine: technology targeting molecular mechanisms for therapy](#). *Journal of internal medicine*, 282(1).
- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. [GloVe: Global vectors for word representation](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.
- Renaud Richardet, Jean-Cédric Chappelier, Martin Telefont, and Sean Hill. 2015. [Large-scale extraction of brain connectivity from the neuroscientific literature](#). *Bioinformatics*, 31(10):1640–1647.
- Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. [Neural machine translation of rare words with subword units](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1715–1725, Berlin, Germany. Association for Computational Linguistics.
- S. Standring and H. Gray. 2008. *Gray's anatomy: The anatomical basis of clinical practice*. Churchill Livingstone/Elsevier, Edinburgh.
- Yuhao Zhang, Peng Qi, and Christopher D. Manning. 2018. [Graph convolution over pruned dependency trees improves relation extraction](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2205–2215, Brussels, Belgium. Association for Computational Linguistics.
- Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, and Christopher D. Manning. 2017. [Position-aware attention and supervised data improve slot filling](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 35–45, Copenhagen, Denmark. Association for Computational Linguistics.