Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230271

Classifiers of Medical Eponymy in Scientific Texts

Dennis TODDENROTH[1]
Chair of Medical Informatics, University Erlangen-Nuremberg, Germany

Abstract. Many concepts in the medical literature are named after persons. Frequent ambiguities and spelling varieties, however, complicate the automatic recognition of such eponyms with natural language processing (NLP) tools. Recently developed methods include word vectors and transformer models that incorporate context information into the downstream layers of a neural network architecture. To evaluate these models for classifying medical eponymy, we label eponyms and counterexamples mentioned in a convenience sample of 1,079 Pubmed abstracts, and fit logistic regression models to the vectors from the first (vocabulary) and last (contextualized) layers of a SciBERT language model. According to the area under sensitivity-specificity curves, models based on contextualized vectors achieved a median performance of 98.0% in held-out phrases. This outperformed models based on vocabulary vectors (95.7%) by a median of 2.3 percentage points. When processing unlabeled inputs, such classifiers appeared to generalize to eponyms that did not appear among any annotations. These findings attest to the effectiveness of developing domain-specific NLP functions based on pre-trained language models, and underline the utility of context information for classifying potential eponyms.

Keywords. Entity classification, medical eponymy, transformer models.

1. Introduction

Concepts that are named after persons are a frequent characteristic of medical terminology and its textual manifestations. Such eponyms may refer to a variety of notions such as diseases, diagnostic signs, therapeutic interventions, or anatomical parts. While most medical eponyms are named after pioneering researchers (such as *Alzheimer disease*), some also originate from affected patients (*Lou Gehrig disease*) or from historical or mythological characters (*ceasarean section*). Eponyms typically resonate with scientific achievement, although their use has also been criticized for misrepresenting academic merit as well as for lacking conceptual accuracy [1].

While existing clinical and scientific texts continue to feature numerous eponyms, however, methods from computational linguistics or natural language processing (NLP) will likely encounter these peculiar phenomena. In contrast to other parts of medical terminology, eponyms are not assembled from semantic elements that provide clues to their medical meaning. Human readers who are not acquainted with a particular condition such as *Addison disease* cannot infer the involved pathophysiology from the eponym; its composed synonym *primary adrenal insufficiency*, on the other hand, is more descriptive in this regard.

[1] Corresponding Author: Dennis TODDENROTH, E-mail: dennis.toddenroth@fau.de.

Other properties like misspellings (*Fischer's exact test*) complicate the automatic detection and analysis of eponyms in conventional biomedical texts. Many clinical terms are named after several persons (*Stevens-Johnson syndrome*), while a few eminent researchers such as Harvey Cushing managed to spawn several concepts. Some terms can occur as eponyms or as non-eponyms, as in a *fractionated dosage of 50 gray* in contrast to *gray matter volume*. Near-homonyms such as *Wegner* and *Wegener* can be misused in patient records, which has motivated the development of preventative clinical decision support functions [2]. The international origin of many medical eponyms can introduce characters that are unconventional in standard English, and inconsistent Anglicization may entail variants such as *Bekhterev* and *Bechterew*. Gradual absorption into parlance may involve a loss of capitalization, as well as inflections and compositions such as *fallopian tube* and *rickettsiosis*.

Previous research has attempted to systematically collect medical eponyms, for example in order to trace their usage, including by automatically expanding search queries with variations from a curated eponym list [3]. Recent applications of advanced NLP approaches have increasingly affected knowledge synthesis from the scientific literature [4]. Such modern NLP methods include word vectors and transformer models that harness high-dimensional numeric representations derived from co-occurrence patterns trained in large textual corpora. This research studies the utility of word vectors and transformer models for recognizing medical eponymy by training and evaluating classifiers in excerpts from the medical literature.

2. Methods

Recent language models represent textual subsequences (tokens) as high-dimensional numeric vectors, or word embeddings. First-layer vocabulary vectors are thereby consistently mapped to the same locations in vector space, regardless of context. Transformer models also use positional encoding and a so-called attention mechanism to incorporate weighted information from surrounding tokens into the downstream vectors of a multilayer neural network. While vocabulary vectors of homonyms as in *Down-regulated gene functions* and *adults with Down syndrome* do not depend on preceding or subsequent tokens, contextual information embedded in the hidden layers of a transformer model thus promises to potentially improve the discrimination of ambiguous eponyms.

The following evaluation considers eponymy classifiers based on a domain-specific language model in a convenience sample of abstracts downloaded from Pubmed. Annotations were defined as either eponyms or non-eponyms using the *Brat Rapid Annotation Tool* [5]. Figure 1 illustrates the annotation procedure as well as the described context-dependent disaggregation of hidden-layer vectors. To improve the efficiency of the annotation process, abstracts were processed in five batches, and candidate labels were automatically pre-annotated with increasingly refined classifier versions. All computed candidate pre-annotations were then manually reviewed and labeled as either eponyms or non-eponyms. Initial pre-annotations were based on the cosine similarity between candidate phrases and average vocabulary vectors from an initial set of eponyms, specified as a plain list of names. Meaningful spatial relations between vocabulary vectors imply that even such a simple approach may potentially generalize to other eponyms beyond the seeded list.

Subsequent batches were pre-annotated by fitting a logistic regression model to contextualized vectors from the last hidden layer of the SciBERT language model, using the *scivocab_uncased* configuration; SciBERT has been pre-trained in a domain-specific corpus, and uses vectors with 768 dimensions [6]. Previous research suggests that models trained in a pertinent biomedical corpus may lead to better performance in domain-specific downstream tasks [7], including the disambiguation of homonyms [8]. Since the SciBERT tokenizer yields a sub-word granularity that disaggregates many eponyms (as in *kl#ats#kin*), word-related vectors were computed by averaging token vectors. Eponym annotations included entire compositions such as *Salmonella* when forming single words, while noun phrases such as *Spearman correlation* were labelled discriminately. Since the attention mechanism is limited to sequences with a maximum of 512 tokens, processing verbose abstracts required that inputs were first separated into a set of chunks, and that outputs were later re-aggregated.

The refined classifier, which potentially discriminates homonyms based on their contexts, was also applied for a plausibility check whereby annotations that appeared to conflict with in-sample predictions were manually revised. The subsequent performance evaluation considered aggregated annotations from all batches as units of observations, and calculated the area under sensitivity-specificity curves in held-out partitions of 100 bootstrapping subsamples. Scripts for computing SciBERT vectors were implemented in Python 3.8, and invoked the language model via the *transformers* package from huggingface.co [9]; regression models were trained via the *glm()* function that ships with of R 3.6. The author, who is trained as a physician, defined and revised all labels. Annotations are available at https://github.com/dtoddenroth/medicaleponyms/.

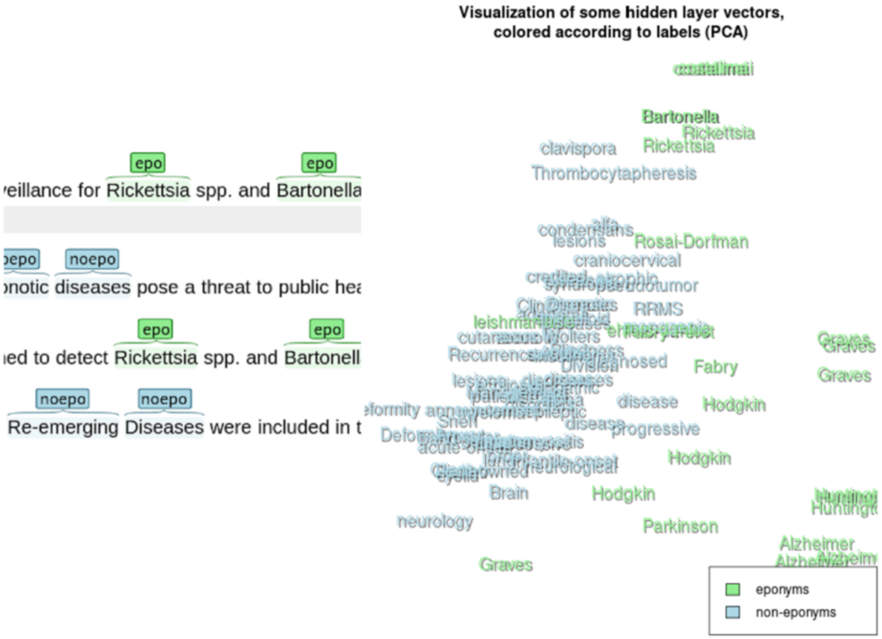

Figure 1. Screenshot of the brat tool during annotation. Candidate sequences are labeled as either eponyms or non-eponyms (left). Visualization of a principal component analysis (PCA) of hidden-layer vectors for exemplary eponyms and counterexamples. While this image shows only some of the information contained in all dimensions, we see moderate label disaggregation. Note that some eponyms appear repeated in different locations, which could allow the context-dependent disambiguation of homonyms (right).

3. Results

The described annotation procedure labelled 1,582 of 13,659 annotations in 1,079 Pubmed abstracts as eponyms (11.6%), which amounts to an average of 1.47 eponyms and 12.7 annotations per abstract. Annotated eponyms included 341 different words, while the three most frequent ones were *Fabry* (227x), *Alzheimer* (148x), and *Parkinson* (81x).

Figure 2 summarizes essential observations from the evaluation of the transformer-based eponymy models. In 100 bootstrap repetitions, logistic regression models trained on first-layer (vocabulary) vectors achieved a median area under the sensitivity-specificity curve of 95.7% (interquartile range 95.4% - 96.1%). Comparable models trained on contextualized vectors from the last hidden layer achieved a median area of 98.0% (97.7% - 98.2%), thus outperforming first-layer models by a median of 2.3 percentage points.

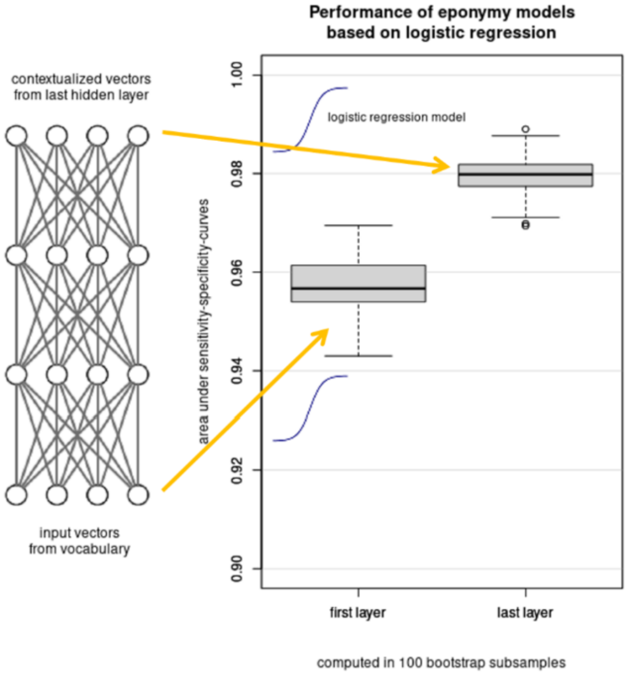

Figure 2. Distribution of the observed performance of eponymy models trained on vocabulary vectors and on contextualized hidden-layer vectors, each computed from 100 bootstrap subsamples. Note that shown y-axis is restricted to areas under sensitivity-specificity curves between 90% and 100%.

During the described batch-wise annotation procedure it also became apparent that intermediate classifier versions successfully generalized to reasonable pre-annotations that did not appear among the previous labels, including *Bland*, *Altman*, *Hounsfield*, and *Kalman*. These anecdotal observations also underline the general capability of the approach.

4. Discussion

The observed performance of the evaluated eponymy classifiers attests to the effectiveness of fitting *'minimal task-specific neural architectures'* to contextualized embeddings [6]. Such eponymy classifiers may be applied to automatically expand previous analyses of their use in literature over time [3]. Note that currently only some eponyms are represented in *Medical Subject Headings* (MeSH), which in Pubmed are assigned to indexed publications. If a more comprehensive set of eponyms such as *Bosworth fracture* could be systematically mapped to explanations like *distal fibula fracture with posterior dislocation of the proximal fragment*, this collection of translations might become useful for improving the reach or precision of pertinent queries to medical literature databases.

As a limitation, the evaluated classifier was restricted to deciding whether a given token sequence constitutes an eponym, and cannot yet properly delimit multi-word entities such as *Bland-Altman*. The superior recognition observed with the contextualized vectors, however, indicates the potential utility of incorporating information from phrases around candidate eponyms, and could also be instrumental for detecting the spans of multi-word eponyms. Previous research has considered various patterns that are typical in the medical literature, including hypernyms [10] and frequent abbreviations [11]. Since annotated datasets could be increasingly valuable for exploring alternative NLP methods in these settings, we hope that the distributed eponym labels might likewise become useful for further experiments.

References

[1] Woywodt A and Matteson E. Should eponyms be abandoned? BMJ. 2007 Sep 1;335(7617):424, doi: 10.1136/bmj.39308.342639.AD.

[2] Baskaran LN, Greco PJ, Kaelber DC. Case report medical eponyms: an applied clinical informatics opportunity. Appl Clin Inform. 2012 Sep 19;3(3):349-55. doi: 10.4338/ACI-2012-05-CR-0019.

[3] He L, Cornish TC, Kricka LJ, Vandergriff TW, Yancey K, Nguyen K, Park JY. Trends in dermatology eponyms. JAAD Int. 2022 Apr 18;7:137-143. doi: 10.1016/j.jdin.2022.03.006.

[4] Wang K and Herr I. Machine-Learning-Based Bibliometric Analysis of Pancreatic Cancer Research Over the Past 25 Years. Front Oncol. 2022. doi: 10.3389/fonc.2022.832385.

[5] Stenetorp P, Pyysalo S, Topic G, Ohta T, Ananiadou S, Tsujii J. brat: a Web-based Tool for NLP-Assisted Text Annotation. Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, pages 102–107, 2012.

[6] Beltagy I, Lo K, Cohan A. SciBERT: A Pretrained Language Model for Scientific Text. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on NLP (EMNLP-IJCNLP). https://www.aclweb.org/anthology/D19-1371.

[7] Wang Y, Liu S, Afzal N, Rastegar-Mojarad M, Wang L, Shen F, Kingsbury P, Liu H. A comparison of word embeddings for the biomedical natural language processing. J Biomed Inform. 2018 Nov;87:12-20. doi: 10.1016/j.jbi.2018.09.008.

[8] Toddenroth D. Evaluation of Domain-Specific Word Vectors for Biomedical Word Sense Disambiguation. Stud Health Technol Inform. 2022 May 16;292:23-27. doi: 10.3233/SHTI220314.

[9] Wolf T, Debut L, Sanh V, Chaumond J, Delangue, C.Moi A et al. Transformers: State-of-the-Art Natural Language Processing. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. 2020. https://www.aclweb.org/anthology/2020.emnlp-demos.6.

[10] Hearst MA. Automatic acquisition of hyponyms from large text corpora. Proceedings of the 14th conference on Computational linguistics. 1992. doi: 10.3115/992133.992154.

[11] Grossman Liu L, Grossman RH, Mitchell EG, Weng C, Natarajan K, Hripcsak G, Vawdrey DK. A deep database of medical abbreviations and acronyms for natural language processing. Sci Data. 2021 Jun 2;8(1):149. doi: 10.1038/s41597-021-00929-4.