

Named Entity Recognition in Functional Neuroimaging Literature

Asma Ben Abacha
National Library of Medicine
Bethesda, MD, USA
asma.benabacha@nih.gov

Alba G. Seco de Herrera
University of Essex
Colchester, UK
alba.garcia@essex.ac.uk

Ke Wang
Swarthmore College
Swarthmore, Pennsylvania, USA
kwang2@swarthmore.edu

L. Rodney Long
National Library of Medicine
Bethesda, MD, USA
long@nlm.nih.gov

Sameer Antani
Senior Member IEEE
National Library of Medicine
Bethesda, MD, USA
antani@nlm.nih.gov

Dina Demner-Fushman
National Library of Medicine
Bethesda, MD, USA
ddemner@mail.nih.gov

Abstract—Human neuroimaging research aims to find mappings between brain activity and broad cognitive states. In particular, Functional Magnetic Resonance Imaging (fMRI) allows collecting information about activity in the brain in a non-invasive way. In this paper, we tackle the task of linking brain activity information from fMRI data with named entities expressed in functional neuroimaging literature. For the automatic extraction of those links, we focus on Named Entity Recognition (NER) and compare different methods to recognize relevant entities from fMRI literature. We selected 15 entity categories to describe cognitive states, anatomical areas, stimuli and responses. To cope with the lack of relevant training data, we proposed rule-based methods relying on noun-phrase detection and filtering. We also developed machine learning methods based on Conditional Random Fields (CRF) with morpho-syntactic and semantic features. We constructed a gold standard corpus to evaluate these different NER methods. A comparison of the obtained F1 scores showed that the proposed approaches significantly outperform three state-of-the-art methods in open and specific domains with a best result of 78.79% F1 score in exact span evaluation and 98.40% F1 in inexact span evaluation.

Keywords—Named Entity Recognition, Functional Magnetic Resonance Imaging, Manual Annotation, Machine Learning, Conditional Random Fields, Rule-based methods

I. INTRODUCTION

Creating a detailed map of brain function [1] requires an effective decoding of cognitive states from patterns of brain activity [2]. Functional Magnetic Resonance Imaging (fMRI) has the capability of mapping brain activity to cognitive states based on changes in blood oxygenation accompanying neural activation [3]. Around 65% of fMRI research study the properties of anatomical brain regions [4] to explore functional localization, cognitive anatomy, or brain structures. Therefore, the textual information contained in the related publications forms a valuable resource for decoding cognitive states. For instance, Nielse et al. [5] proposed a method to find association rules between terms extracted from neuroimaging studies and Talairach coordinates using a bag-of-words representation. Hsiao et al. [6] also used fMRI literature with the

Unified Medical Language System[®] (UMLS[®]) to construct a hierarchical concept-based dictionary of brain functions. A named entity recognition approach was used in Neurosynth [7] which extracts named entities from neuroimaging publications and generates probabilistic mappings between cognitive and neural states based on entity frequencies. Other systems include WhiteText [8] which extracts brain regions and their connections from neuroscience literature using co-occurrence methods, and BrainSCANr [9], which provides a dictionary of brain region names, cognitive and behavioral functions and diseases, and calculates the co-occurrence between two terms in the scientific literature. linkRbrain [10] identifies cognitive task expressions by using the CorText platform followed by a manual selection. In this paper, we focus on recognizing named entities that can be relevant to describe relationships between brain activity and a large number of broad cognitive states. We propose automatic NER methods from functional neuroimaging literature and we compare them to three baselines.

II. NAMED ENTITY RECOGNITION METHODS

NER is a fundamental step in natural language understanding aiming to identify phrases referring to entities from predefined categories [11]–[13]. The selection of a particular NER method depends on the domain (general vs. specific), the difficulty of the task (e.g. the definition of named entities), and the availability of training data or predefined rules. For NER in functional neuroimaging articles, there are no readily available training data, and no predefined categories or rules for named entities. Therefore, we defined relevant entity categories and developed rule-based methods as well as machine-learning methods trained on manually annotated abstracts.

A. Entity Categories

Neuroscientists are interested in brain responses to specific experimental tasks, the areas of the brain in which these

responses occur, and the brain functions implied by these responses [6]. We defined 15 named entity categories following this fundamental observation: “Gross brain anatomy”, “Functional neuroanatomy”, “Brain function”, “Body anatomy”, “Body function”, “Medical problem” and “Sensory stimuli or response” which includes eight sub-categories: “Gustation”, “Visual”, “Emotional”, “Olfactory”, “Auditory”, “Somatosensory”, “Abstract”, and “Other”.

B. Rule-based Methods

We used a three-pronged method for NER based on noun-phrase detection and word filtering.

1) *Noun Phrase Chunking*: Noun phrase detection consists of finding noun phrases in a given text. We chose to use Tree-Tagger [14] as our first chunker [12] and Stanford Parser [15].

2) *Rules for Entity Boundaries*: Annotation guidelines differ according to the task (e.g. the inclusion or not of articles, adjectives, and numbers). We defined a set of rules to change the boundaries of noun phrases using the morphosyntactic categories of the words (e.g. determiner, adjective). We automatically removed the determiners (e.g. an, those, no, each), personal and possessive pronouns (e.g. her, their), cardinal numbers and specific adjectives (e.g. most, same) from the noun phrases. Additional rules are defined for entity length (between 2 and 55 characters) and coordinating conjunctions.

3) *Word Filtering*: Stopword removal is an essential step to filter irrelevant words. We use a list of 723 common stopwords in information retrieval (e.g. about, normally, useful). As filtering domain-related words is also important, we constructed and manually validated a second list of 3,405 functional stopwords related to neuroimaging using the most frequent words in a collection of 11,406 fMRI articles (e.g. fMRI, results, activation, patients, regions, activity, findings).

C. Machine Learning Methods

We built a CRF classifier using sklearn-crfsuite¹. Training was performed on 91 abstracts randomly selected from the Neurosynth collection and manually annotated using Brat². We selected the following list of features for machine learning:

1) *Word and Morpho-syntactic Features*: For each word: (i) the word itself, lemma of the word, part of speech tag (POS), (ii) prefixes and suffixes, and (ii) presence in a stopword list. The lemma and POS tag are extracted using NLTK³. Information on words immediately before and after this particular word in a sentence are included as well: (i) the previous word and its POS; and (ii) the following word and its POS.

2) *Semantic Features*: We used 3 semantic categories and generated a semantic feature indicating whether or not the word belongs to the term list of one of the categories: (i) Medical Problem list of 7,062 terms extracted from Biomedical Entity Network⁴, Wikipedia (list of diseases⁵, cancer types⁶)

and Health on the Net (rare diseases⁷); (ii) Human Body Anatomy list of 163 terms from Human Body Vocabulary⁸; and (iii) Gross Brain Anatomy, with a list of 554 terms using available resources such as Allen Brain Atlas⁹, the Whole Brain Atlas¹⁰ and Neuroanatomy¹¹.

D. Baseline Methods

We selected three methods and tools as baselines for NER from functional neuroimaging articles.

1) *Neurosynth*: a widely used platform for automatically synthesizing the results of different neuroimaging studies [7]. We used the Neurosynth lexicon containing 3,169 terms extracted from the abstracts (version of July 2015) to annotate our collection.

2) *DBPedia-KODA*: one of the best performing entity linking tools [16]. We used the DBpedia implementation of KODA to annotate our collection.

3) *MeSH*[®]: a controlled and structured vocabulary of medical topics provided and maintained by the U.S. National Library of Medicine (NLM). We automatically extracted MeSH[®] terms from XML articles and used them as named entities.

III. EVALUATION METHODOLOGY

To create a gold standard for evaluating the performance of the NER methods, a manual annotation was carried out by two annotators on abstracts of 52 neuroimaging articles randomly selected from the Neurosynth collection. We also developed guidelines for the annotation process performed using Brat. The resulting corpus contains 864 annotated named entities. We computed the Inter-Annotator Agreement (IAA) using the Bio-SCoRes assessment code¹² [17]. Precision (P), Recall (R) and F1 were computed for exact and overlapping span matches. Table I shows the global IAA results.

TABLE I
OVERALL INTER-ANNOTATOR AGREEMENT.

Eval. Criteria	Exact Span			Inexact Span		
	P	R	F1	P	R	F1
Span and Category	37.62	34.82	36.16	51.27	47.46	49.29
Entity Span only	57.22	52.26	54.63	81.02	78.43	79.70

IAA is computed based on the span and category of annotated entities, then based on the entity span only. Both exact and inexact boundary matching are considered for the evaluation. Inexact matching is a flexible matching between two extracted named entities expressing that they are overlapping or equal. Table II shows the IAA for each category. The best IAA is obtained in the “gross brain anatomy” category which indeed has the highest number of named entities in the corpus. Other categories are more difficult to annotate because of their ambiguity (e.g. “somatosensory stimuli or response”).

¹<https://sklearn-crfsuite.readthedocs.io/en/latest/>

²<http://brat.nlpab.org/>

³<http://www.nltk.org/>

⁴<http://crn.vistainformatics.com/>

⁵https://en.wikipedia.org/wiki/Lists_of_diseases/

⁶https://en.wikipedia.org/wiki/List_of_cancer_types/

⁷<http://www.hon.ch/HONselect/RareDiseases/index.html/>

⁸<http://www.enchantedlearning.com/wordlist/body.shtml/>

⁹<http://www.brain-map.org/>

¹⁰<http://www.med.harvard.edu/aanlib/>

¹¹<https://brinaacn.org/neuroanatomy/>

¹²<https://github.com/kilicogluh/Bio-SCoRes/>

Finally, a reconciliation was carried out to correct errors in the annotated abstracts of both annotators and create a unique gold standard.

TABLE II
INTER-ANNOTATOR AGREEMENT FOR EACH ENTITY CATEGORY.

Entity (E) Category	#E	Exact Matching			Inexact Matching		
		P	R	F1	P	R	F1
Gross brain anatomy	258	54.5	61.2	57.6	71.5	80.4	75.7
Functional neuroanatomy	41	45.5	35.7	40.0	60.6	47.6	53.3
Brain function	84	30.4	17.1	21.9	32.6	18.3	23.4
Body anatomy	18	28.6	42.9	34.3	33.3	50.0	40.0
Body function	38	37.5	25.7	30.5	41.7	28.6	33.9
Medical problem	53	56.8	44.7	50.0	73.0	57.5	64.3
Stimuli/responses	Gustation	3	0	0	0	0	0
	Visual	87	26.1	22.2	24.0	50.7	43.2
	Olfactory	9	0	0	0	0	0
	Auditory	51	40.0	25.0	30.8	56.7	35.4
	Somatosensory	26	25.0	8.7	12.9	37.5	13.04
	Emotional	61	18.8	29.6	23.0	25.9	40.7
	Abstract	89	19.4	16.9	18.1	30.6	26.5
	Other	55	2.6	2.1	2.3	10.3	8.3

IV. EXPERIMENTAL RESULTS

TABLE III
NER RESULTS. (ML) MACHINE-LEARNING, (SF) SEMANTIC FEATURES, (R) RULE-BASED, (B) BASELINES, (M) MANUAL ANNOTATION.

Method		Exact Matching			Inexact Matching		
		P	R	F1	P	R	F1
ML	CRF+SF	81.1	76.6	78.8	99.6	97.1	98.4
	CRF	69.4	32.2	44.0	91.9	46.1	61.4
R	TreeTagger+R	30.7	55.7	39.6	50.3	86.4	63.6
	StanfordParser+R	23.9	61.4	34.4	51.0	93.4	66.0
B	KODA	13.0	51.2	20.7	39.2	92.9	55.2
	MeSH [®]	4.4	3.9	4.1	12.4	11.8	12.1
	Neurosynth	3.1	19.7	5.3	32.1	98.1	48.4
M	Annotator (A)	67.0	56.6	61.4	90.0	80.3	84.6
	Annotator (B)	96.6	89.3	92.8	99.5	92.66	95.4

Table III presents the results of our automatic methods, baseline methods and manual annotations (we do not consider the entities categories in this evaluation). Our automatic methods (Machine-learning and Rule-based) obtained significantly better results than the baseline systems. The open-domain tool KODA achieved good recall in inexact span but precision is poor because of the specificity of the domain. The Neurosynth method obtained good recall in inexact span but it does not fit the goal of extracting relevant named entities with correct boundaries. Neurosynth recognizes general words as entities (e.g. using, repeat, asked), entities with inexact boundaries (e.g. amygdala insula) and also irrelevant entities for this specific domain (e.g. magnetic resonance, brains). MeSH[®] terms also were not appropriate for the task.

V. CONCLUSION

We studied the recognition of named entities for the automated understanding of brain anatomy and the brain cognitive functions expressed in publications related to fMRI experiments. Results showed that the proposed NER methods outperform state-of-the-art methods in open and specific domains. Using CRF and relevant semantic features, we achieve 78.79% F1 score in exact evaluation and 98.40% F1 in inexact evaluation. Since there is no gold standard, a human annotation

effort was involved to construct a gold standard corpus for the task which will be made available. Our final goal is to develop a robust tool to map fMRI brain activations with relevant entities and thus, decode cognitive states from brain activity. This capability will allow using and integrating information contained in neuroscience publications at a large scale.

ACKNOWLEDGMENTS

This research was supported by the Intramural Research Program of the National Institutes of Health, National Library of Medicine, and Lister Hill National Center for Biomedical Communications. We thank Yassine Mrabet for providing KODA annotations and Halil Kilicoglu for his support with IAA assessment.

REFERENCES

- [1] T. Dimitrova, "Brain decoding: Modern day cartographers aim to map the human brain," 2015.
- [2] R. Adolphs, "The unsolved problems of neuroscience," *Trends in Cognitive Sciences*, vol. 19, no. 4, pp. 173–175, 2015.
- [3] R. B. Buxton, *Introduction to functional magnetic resonance imaging: principles and techniques*. Cambridge university press, 2009.
- [4] N. K. Logothetis, "What we can do and what we cannot do with fMRI," *Nature*, vol. 453, no. 7197, pp. 869–878, 2008.
- [5] F. Å. Nielsen, L. K. Hansen, and D. Balslev, "Mining for associations between text and brain activation in a functional neuroimaging database," *Neuroinformatics*, vol. 2, no. 4, pp. 369–379, 2004.
- [6] M.-Y. Hsiao, C.-C. Chen, and J.-H. Chen, "Using UMLS to construct a generalized hierarchical concept-based dictionary of brain functions for information extraction from the fMRI literature," *Journal of biomedical informatics*, vol. 42, no. 5, pp. 912–922, 2009.
- [7] T. Yarkoni, R. A. Poldrack, T. E. Nichols, D. C. Van Essen, and T. D. Wager, "Large-scale automated synthesis of human functional neuroimaging data," *Nature methods*, vol. 8, no. 8, pp. 665–670, 2011.
- [8] L. French, S. Lane, L. Xu, C. Siu, C. Kwok, Y. Chen, C. Krebs, and P. Pavlidis, "Application and evaluation of automated methods to extract neuroanatomical connectivity statements from free text," *Bioinformatics*, vol. 28, no. 22, pp. 2963–2970, 2012.
- [9] J. B. Voytek and B. Voytek, "Automated cognome construction and semi-automated hypothesis generation," *Journal of neuroscience methods*, vol. 208, no. 1, pp. 92–100, 2012.
- [10] S. Mesmoudi, M. Rodic, C. Cioli, J.-P. Cointet, T. Yarkoni, and Y. Burnod, "LinkRbrain: Multi-scale data integrator of the brain," *Journal of neuroscience methods*, vol. 241, pp. 44–52, 2015.
- [11] W. W. Chapman, W. Bridewell, P. Hanbury, G. F. Cooper, and B. G. Buchanan, "A simple algorithm for identifying negated findings and diseases in discharge summaries," vol. 34, no. 5, pp. 301–310, 2001.
- [12] A. Ben Abacha and P. Zweigenbaum, "Medical entity recognition: A comparison of semantic and statistical methods," in *Proceedings of the 10th ACL workshop on Biomedical Natural Language Processing (BioNLP)*. Portland, Oregon, USA: Association for Computational Linguistics, 2011, pp. 56–64. [Online]. Available: <http://www.aclweb.org/anthology/W11-0207>
- [13] D. Nadeau and S. Sekine, "A survey of named entity recognition and classification," *Linguisticae Investigationes*, vol. 30, no. 1, pp. 3–26, 2007.
- [14] H. Schmid, "Probabilistic part-of-speech tagging using decision trees," in *International Conference on New Methods in Language Processing*, Manchester, UK, 1994, pp. 44–49.
- [15] D. Chen and C. Manning, "A fast and accurate dependency parser using neural networks," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, October 2014, pp. 740–750. [Online]. Available: <http://www.aclweb.org/anthology/D14-1082>
- [16] Y. Mrabet, C. Gardent, M. Foulonneau, E. Simperl, and E. Ras, "Towards knowledge-driven annotation," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA.*, 2015, pp. 2425–2431. [Online]. Available: <http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9309>
- [17] H. Kilicoglu and D. Demner-Fushman, "Bio-SCoRes: A smorgasbord architecture for coreference resolution in biomedical text," *PLoS one*, vol. 11, no. 3, p. e0148538, 2016.