MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231054

Identifying Mentions of Pain in Mental Health Records Text: A Natural Language Processing Approach

Jaya CHATURVEDI[a,1] , Sumithra VELUPILLAI[a], Robert STEWART[a,b,c] and Angus ROBERTS[a, b]
[a] *Institute of Psychiatry, Psychology and Neurosciences, King's College London*
[b] *Health Data Research, UK*
[c] *South London and Maudsley Biomedical Research Centre, London, United Kingdom*
ORCiD ID: Jaya Chaturvedi https://orcid.org/0000-0002-6359-9853

Abstract. Pain is a common reason for accessing healthcare resources and is a growing area of research, especially in its overlap with mental health. Mental health electronic health records are a good data source to study this overlap. However, much information on pain is held in the free text of these records, where mentions of pain present a unique natural language processing problem due to its ambiguous nature. This project uses data from an anonymised mental health electronic health records database. A machine learning based classification algorithm is trained to classify sentences as discussing patient pain or not. This will facilitate the extraction of relevant pain information from large databases. 1,985 documents were manually triple-annotated for creation of gold standard training data, which was used to train four classification algorithms. The best performing model achieved an F1-score of 0.98 (95% CI 0.98-0.99).

Keywords. Natural language processing, electronic health records, pain, mental health, transformers.

1. Introduction

Pain is defined as an unpleasant sensory and emotional experience, and is influenced by a variety of biological, psychological, and social factors [1]. Pain is a common reason for people to access healthcare facilities, thereby making electronic health records (EHR) a potential source for information on pain [2].

EHRs are longitudinal compilations of electronic data pertaining to a person's medical history or healthcare [3]. They have been increasingly used in research as they provide the opportunity to explore patient symptoms and findings from structured and unstructured fields. Since pain is not well recorded in these structured fields, it may help to supplement this information with data from unstructured clinical text [4].

A commonly used machine learning based NLP approach is text classification, in which labels are assigned to units of text (sentences/paragraphs/documents) [5]. Commonly used classification algorithms include Support Vector Machines (SVM) [6] and K-Nearest Neighbours (KNN) [7]. Recent state of the art approaches use embedding

[1] Corresponding Author: Jaya Chaturvedi, email: jaya.1.chaturvedi@kcl.ac.uk.

models and transformer-based neural network architectures [8], such as the bi-directional encoder representations of BERT [9]. Many healthcare domain related models have emerged, such as UmlsBERT [10] and SAPBERT [11] which were developed after recognition of the need for specialized models due to linguistic differences between general and biomedical text [12].

This paper describes the methods undertaken to develop an NLP application for a sentence-level classification of mentions of physical pain within clinical text. Two BERT models were trained - BERT_base and SAPBERT - and compared to two conventional models - SVM and KNN. To the best of our knowledge, such extraction of information about pain from mental health clinical text using NLP has not been done.

2. Methods

Data Source

An anonymised version of EHR data from The South London and Maudsley NHS Foundation Trust (SLaM), one of the largest mental healthcare organizations in Europe, is stored in the Clinical Record Interactive Search (CRIS) database [13]. The infrastructure of CRIS has been described in detail with an overview of the cohort profile [14].

Ethics and Data Access

Ethics approval for CRIS has been granted by (Oxford C Research Ethics Committee, reference 18/SC/0372). Research projects that use the CRIS database are reviewed and approved by a patient-led oversight committee [15]. An opt-out model is in place for service users and is advertised in all publicity material and initiatives.

Data Extraction

Pain can be described in numerous ways, using a variety of terms. To help identify which documents in CRIS might be discussing pain, a lexicon of such pain terms was developed from a combination of pain-related terms extracted from the literature, biomedical ontologies, and additional similar terms from word embedding models [16]. Documents containing pain terms were extracted using SQL. No time or diagnosis filter was applied to the extraction.

Annotation Task

Extracted documents were used to create a corpus of text discussing patient pain by labelling, i.e., annotating, spans of text as being about pain or not. Each span consisted of 200 characters before and after a pain-related term. First, a set of annotation guidelines were developed to provide rules defining when a sentence should be considered as discussing pain. Next, terms from the pain lexicon were highlighted in the extracted documents. Three medical student annotators read through the extracted documents considering these spans of text containing the previously highlighted pain terms. Annotators labelled each span with one of three labels: *relevant* i.e., referring to physical pain experienced by the patient; *not relevant* i.e., mentions not related to pain, not related to the patient or hypothetical and metaphorical mentions; and *negated* i.e., absence of pain. The annotation tool used for this was MedCAT [17].

NLP application

The annotations were split into train/test/validation sets at a proportion of 80/10/10 respectively. Four different models were trained, as detailed in Table 1. The parameters for the BERT models were chosen based on the recommendations made in [9] and models were checked for overfitting.

Table 1. Model specifications.

Model	Tokenizer	Pre-processing	Other Parameters
1.Support Vector Machine 2.K-Nearest Neighbour	NLTK	Lowercase, stopword, white space and punctuation removal, lemmatize and tokenize	Tf-Idf vectorizer Default parameters from sklearn
3. BERT_base	bert_base_uncased	Tokenize Prepend sentence with special token [CLS] and append with special token	Epochs: 3 Batch size: 16 Optimizer: AdamW, learning rate 3e-5
4. SAPBERT	cambridgeltl/SapBERT-from-PubMedBERT-fulltext	[SEP] Pad and truncate sentence to max length 105 (default is 511)	Epochs: 4 Batch size: 16 Optimizer: AdamW, learning rate 2e-5

3. Results

Data Extraction

A total of 1,985 randomly selected documents from 723 patients were extracted that contained pain related keywords from the lexicon. The most common diagnosis codes for these extracted patients were Mood disorders (ICD10 chapters F30-39) (33% of patients). There was an average of 8 annotations per patient.

Annotations

An inter-annotator agreement of 90% (Cohen's kappa 0.88) was achieved. A total of 5,644 annotations were obtained. 72% of these were marked as relevant, 15% as not-relevant, and 13% as negated. The relevant annotations were labelled as 1. The not-relevant and negated annotations were combined and labelled as 0.

Evaluation of NLP application

K-fold validation was carried out for evaluation of the models, and 95% confidence intervals calculated (Table 2) and BERT models performed better.

Table 2. Evaluation Metrics, including 95% confidence intervals.

Model	Precision	Recall	F1-score (average from 10-fold cross validation)
Support Vector Machine	0.86 (0.83-0.88)	0.98 (0.97-0.99)	0.91 (0.90-0.93)
K-Nearest Neighbour	0.84 (0.81-0.87)	0.91 (0.89-0.93)	0.87 (0.85-0.89)
BERT_base	0.96 (0.94-0.97)	0.98 (0.97-0.99)	0.97 (0.96-0.98)
SAPBERT	0.98 (0.97-0.99)	0.99 (0.98-0.99)	0.98 (0.98-0.99)

Error Analysis

During the annotation process, common disagreements included when an instance could be interpreted as physical or metaphorical, such as "...causing him pain", and hypothetical mentions such as "...she feared the pain" and "?migraine".

The SAPBERT model showed false negatives when there were undecipherable symbols incorporated in the text, as well as misspellings or conjoined words such as "dabdominal pain" and "achespainodd sensations". False positives were instances such as "risk of potential pressure sores".

4. Discussion

The ambiguous nature of pain was highlighted during this project. This highlights the importance of context and the necessity for NLP models to incorporate and consider context during the classification task. This is a strength of transformer-based models such as BERT, which could be why they performed better than SVM/KNN.

Amongst the two BERT models that were trained, SAPBERT, which was pre-trained using a biomedical ontology, UMLS, performed slightly better than BERT_base. There were differences in how each of the BERT models used in this project tokenised words, where SAPBERT was able to tokenise clinical concepts more accurately. This improvement in tokenisation might have impacted and improved the overall performance of the model.

5. Conclusions

The objective of this project was to develop a machine learning based NLP application that can classify mentions of pain within clinical text as relevant or not. BERT models outperformed the other algorithms. This is a novel approach towards extracting information about pain from mental health records, leveraging the unstructured clinical notes to identify patients with relevant mentions of pain, and such cohorts of patients can then further be used in epidemiological and other pain related research with more confidence in the actual occurrence of pain when mentioned in the text.

Acknowledgements

This work uses data provided by patients and collected by the NHS as part of their care and support. The authors are also grateful to Dr Aurelie Mascio for providing access to some of her Python scripts.

References

[1] Terminology. International Association for the Study of Pain. International Association for the Study of Pain (IASP). [cited 2022 Oct 31]. Available from: https://www.iasp-pain.org/resources/terminology/

[2] Motov SM, Khan AN. Problems and barriers of pain management in the emergency department: Are we ever going to get better? J Pain Res. 2008 Dec;2:5-11, doi: 10.2147/JPR.S4324.

[3] Safety I of M (US) C on DS for P, Aspden P, Corrigan JM, Wolcott J, Erickson SM. Key Capabilities of an Electronic Health Record System: Letter Report [Internet]. Patient Safety: Achieving a New Standard for Care. National Academies Press (US); 2004 [cited 2022 Oct 31]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK216105/

[4] Carlson LA, Jeffery MM, Fu S, He H, McCoy RG, Wang Y, Hooten WM, St Sauver J, Liu H, Fan J. Characterizing chronic pain episodes in clinical text at two health care systems: comprehensive annotation and corpus analysis. JMIR Med Inform. 2020 Nov;8(11):e18659, doi: 10.2196/18659.

[5] Minaee S, Kalchbrenner N, Cambria E, Nikzad N, Chenaghlu M, Gao J. Deep learning based text classification: a comprehensive review. ACM computing surveys (CSUR). 2021 Apr;54(3):1-40, doi: 10.1145/3439726.

[6] Wright A, McCoy AB, Henkin S, Kale A, Sittig DF. Use of a support vector machine for categorizing free-text notes: assessment of accuracy across two institutions. J Am Med Inform

[7] Trstenjak B, Mikac S, Donko D. KNN with TF-IDF based framework for text categorization. Procedia Eng. 2014 Jan;69:1356–64, doi: 10.1016/j.proeng.2014.03.129.

[8] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention Is All You Need [Internet]. arXiv; 2017 [cited 2022 Oct 31]. Available from: http://arxiv.org/abs/1706.03762

[9] Devlin J, Chang MW, Lee K, Toutanova K. 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. Minneapolis, Minnesota: Association for Computational Linguistics; 2019. p. 4171–86. Available from: https://www.aclweb.org/anthology/N19-1423

[10] Michalopoulos G, Wang Y, Kaka H, Chen H, Wong A. UmlsBERT: clinical domain knowledge augmentation of contextual embeddings using the unified medical language system metathesaurus. arXiv:201010391.2020 Oct, doi: 10.48550/arXiv.2010.10391

[11] Liu F, Shareghi E, Meng Z, Basaldella M, Collier N. Self-alignment pretraining for biomedical entity representations. arXiv:201011784. 2020 Oct, doi: 10.48550/arXiv.2010.11784.

[12] Alsentzer E, Murphy J, Boag W, Weng WH, Jin D, Naumann T, McDermott MBA. Publicly available clinical BERT embeddings. In: Proceedings of the 2nd Clinical Natural Language Processing Workshop. Minneapolis, Minnesota, USA: Association for Computational Linguistics; 2019. p. 72–8, doi: 10.48550/arXiv.1904.03323.

[13] Stewart R, Soremekun M, Perera G, Broadbent M, Callard F, Denis M, Hotopf M, Thornicroft G, Lovestone S. The south london and maudsley NHS foundation trust biomedical research centre (SLAM BRC) case register: development and descriptive data. BMC Psychiatry. 2009 Aug;9:51, doi: 10.1186/1471-244X-9-51

[14] Perera G, Broadbent M, Callard F, Chang CK, Downs J, Dutta R, Fernandes A, Hayes RD, Henderson M, Jackson R, Jewell A, Kadra G, Little R, Pritchard M, Shetty H, Tulloch A, Stewart R. Cohort profile of the south london and maudsley NHS foundation trust biomedical research centre (SLaM BRC) case register: current status and recent enhancement of an electronic mental health record-derived data resource. BMJ Open. 2016 Mar;6(3):e008721, doi: 10.1136/bmjopen-2015-008721.

[15] Fernandes AC, Cloete D, Broadbent MT, Hayes RD, Chang CK, Jackson RG, Roberts A, Tsang J, Soncul M, Liebscher J, Stewart R, Callard F. Development and evaluation of a de-identification procedure for a case register sourced from mental health electronic records. BMC Med Inform Decis Mak. 2013 Jul;13:71, doi: 10.1186/1472-6947-13-71.

[16] Chaturvedi J, Mascio A, Velupillai SU, Roberts A. Development of a lexicon for pain. Front Digit Health. 2021 Dec;3:778305, doi: 10.3389/fdgth.2021.778305.

[17] Kraljevic Z, Searle T, Shek A, Roguski L, Noor K, Bean D, et al. Multi-domain Clinical Natural Language Processing with MedCAT: the Medical Concept Annotation Toolkit. arXiv:201001165 [cs] [Internet]. 2021 Mar 25 [cited 2022 Feb 26]; Available from: http://arxiv.org/abs/2010.01165