Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 2268–2275
Marseille, 11–16 May 2020

The Language of Brain Signals: Natural Language Processing of Electroencephalography Reports

Ramon Maldonado, Sanda M. Harabagiu
The University of Texas at Dallas
800 W. Campbell Rd., Richardson, TX 75080
{ramon,sanda}@utdallas.edu

Abstract

Brain signals are captured by clinical electroencephalography (EEG) which is an excellent tool for probing neural function. When EEG tests are performed, a textual EEG report is generated by the neurologist to document the findings, thus using language that describes the brain signals and their clinical correlations. Even with the impetus provided by the BRAIN initiative (braininititive.nih.gov), there are no annotations available in texts that use natural language describing the brain activities and their correlations with various pathologies. In this paper we describe an annotation effort carried out on a large corpus of EEG reports, providing examples of EEG-specific and clinically relevant concepts. In addition, we detail our annotation schema for brain signal attributes. We also discuss the resulting annotation of long-distance relations between concepts in EEG reports. By exemplifying a self-attention joint-learning method used to predict concept, attribute and relation annotations in the EEG report corpus, we discuss the promising results of automatic annotations, hoping that our effort will inform the design of novel knowledge capture techniques that will include the language of brain signals.

Keywords: language of the brain, medical concepts, long-distance relations.

1. Introduction

The diagnosis and clinical management of epilepsy as well as the evaluation of other types of brain disorders (Smith, 2005), including encephalopathies, neurological infections, Creutzfeldt-Jacob disease, and even in the evaluation of the progression of Alzheimer's disease are informed by clinical electroencephalographies (EEG). An EEG records electrical activity along the scalp and measures spontaneous electrical activity of the brain. The signals measured along the scalp can be correlated with brain activity. But, as noted in (Beniczky et al., 2013), the complexity of the EEG signal, interpreted and documented in EEG reports, produces inter-observer agreement known to be moderate. As more clinical EEG signals and reports become available, the interpretation of EEG signals can be improved by providing neurologists with results of search for patients that exhibit similar EEG characteristics.

Neurologists that are interested in patients that exhibit certain EEG characteristics can use in their their queries medical concepts that characterize the EEG tests, e.g. EEG activities of the brain, EEG events, or the clinical correlations between brain signals and pathologies of the brain. In order to discover relevant EEG reports, there is a need to automatically identify EEG-specific concepts as well as relations they hold to the clinical correlations of the brain signals. The ability to identify such concepts is hindered by the absence of annotations in medical texts covering EEG-specific information.

In this paper we describe an annotation effort performed on a corpus of 23,000 EEG reports from over 15,000 patients collected over 12 years at the Temple University Hospital (TUH). The corpus is freely available after registration[1]. Moreover, this corpus is, to our knowledge, the only publicly available resource of EEG signals and EEG reports. Because no previous language annotations were performed on any EEG reports, we believe that our effort fills a gap in clinical informatics as well as in clinical language resources - as it is the first effort performed on this form of clinical narratives. The annotations were performed through a combination of deep and active learning, reported in (Maldonado and Harabagiu, 2019). The resulting annotations consists of (1) a gold-standard set, vetted by experts in neurology and natural language processing; and (2) a silver-standard set, produced by a neural learning technique that uses self-attention and benefits from active learning. Both sets of annotations may serve the automatic discovery of connections between abnormal brain signals and the pathologies of the brain. These annotations could also inform knowledge acquisitions methods operating both on Electronic Health Records (EHRs) and medical literature. In our previous work (Goodwin and Harabagiu, 2016), we used these annotations to produce a multi-modal index for patient cohort retrieval.

In summary, the contributions of the work described in this paper is three-fold:

1. Annotations on a corpus of EEG reports, which were designed to convey a written impression of the visual analysis of the brain signals along with its clinical significance. The annotations capture both the EEG-specific concepts as well as general clinical concepts, such as medical problems, treatments and other tests;

2. Annotation of the attributes of brain signals, recognized in EEG activities, which describe the morphology of the signal; the frequency band; the brain background; the recurrence of the brain signals; as well as a set of spacial attributes of the brain signals, including the brain hemisphere were they occur or the more specific brain location, as well as the dispersal of these signals.

3. Annotations of long-distance relations between pairs

[1] https://www.isip.piconepress.com/projects/tuh_eeg/

of EEG-specific and clinically relevant concepts from the same EEG report.

The remainder of the paper is organized as follows. Section 2 presents structure and content of EEG reports while Section 3 details the concepts, attributes and relations that were annotated in the corpus of EEG reports. Section 4 briefly discusses the automatic annotation while Section 5 presents the annotation results. Section 6 summarizes the conclusions.

2. Structure and Content of EEG Reports

The American Clinical Neurophysiology Society Guidelines for writing EEG reports mandates that each EEG report starts with a *clinical history* of the patient including information about the patient's age, gender, current medical conditions (e.g. *"right leg swelling"*), and relevant past medical conditions (e.g. *"asthma"*) followed by a list of *medications* the patient is currently taking (e.g. *"Albuterol"*), described in a separate section. Together, these two initial sections depict *the clinical picture and therapy* of the patient, containing a wealth of medical concepts including *medical problems* (e.g. *"right leg swelling"*), symptoms (e.g. *"facial droop"*), signs (e.g. *"twitching"*) and treatments (e.g. *"Albuterol"*, *"gastrocnemius surgery"*). After the clinical picture and therapy of the patient is established, the *introduction* section of the EEG report describes the techniques used for the current EEG (e.g. *"digital video EEG using standard 10-20 system of electrode placement with one channel of EKG"*), the patient's condition at the time of the record (e.g. *wakefulness*), and possible activating procedures carried out (e.g. *"photic stimulation"*).
The *description* section is the mandatory part of the report, meant to provide a complete and objective description of the EEG, noting all observed *EEG activities* (e.g. *"frontocentral beta"*, *"triphasic waves"*), and *EEG events* (e.g. *"stimulation"*). The *impression* section indicates whether or not the EEG test is abnormal and, if so, lists the abnormalities in decreasing order of importance. These abnormalities are usually describing EEG activities (e.g. *"triphasic waves"*), but can also be EEG Events. Finally, the *clinical correlation* section explains what the EEG findings mean in terms of clinical interpretation, (e.g. *"findings indicative of underlying metabolic encephalopathy"*).
As illustrated in Figure 1, each of the sections of EEG reports mentions multiple forms of medical concepts, which have annotations indicating medical problems [PROB], treatments [TR], tests [TEST], EEG activities [ACT], and EEG events [EV]. In addition, these medical concepts are characterized by various attributes, which we have also annotated. The attributes are defined in Section 3.1. The recognition of medical concepts and their attributes in EEG reports is vital for many applications requiring data-driven representation of EEG-specific knowledge, including decision support systems. However, the identification of the medical concepts in the EEG reports is not sufficient, as these concepts also exhibit clinically-relevant relations between them. Without knowledge of these relations the acquired EEG-specific clinical knowledge would not be represented in a graphical format, but rather a list of concepts mentioned in EEG

CLINICAL HISTORY: 35 year old right handed female with history of [asthma]$_{PROB}$ and no family history of [seizures]$_{PROB}$ or [head trauma]$_{PROB}$. Last year she had [gastrocnemius surgery]$_{TR}$. She presents with [right leg swelling]$_{PROB}$. She recently developed [twitching]$_{PROB}$ and there is a noticeable [facial droop]$_{PROB}$.
MEDICATIONS: [Albuterol]$_{TR}$, many others.
INTRODUCTION: [Digital video EEG]$_{TEST}$ is performed in lab using standard 10-20 electrode placement system with anterior temporal electrode and [EKG]$_{TEST}$ electrodes. Wakefulness and sleep stage 1, stage 2 were recorded. Aggravated procedures with [hyperventilation]$_{EV}$ and [photic stimulation]$_{EV}$ were performed.
DESCRIPTION OF THE RECORD: The record opens to a well formed posterior dominant rhythm at 9Hz with 2260 microvolts amplitude and normal frontocentral beta. There are frontally predominant, relatively synchronous [triphasic waves]$_{ACT}$ seen throughout the record. [Stimulation]$_{EV}$ of the patient produces cessation of the [triphasic waves]$_{ACT}$. [Focal slowing]$_{ACT}$ was seen at T3 at 2-4 Hz at times in runs of 2. not clearly monomorphic.
IMPRESSION: Abnormal EEG due to:
1. [Triphasic waves]$_{ACT}$.
2. [Slowing]$_{ACT}$.
CLINICAL CORRELATION: No [seizures]$_{PROB}$ were recorded. The [triphasic waves]$_{ACT}$ are indicative of underlying [metabolic encephalopathy]$_{PROB}$ including [hepatic encephalopathy]$_{PROB}$.

Figure 1: Synthetic Example EEG Report.

reports. Therefore, several clinically relevant relations between concepts have also been identified and annotated in the corpus. It is important to note that the relations we have annotated are long-distance relations, connecting often concepts that are mentioned in different sections of the EEG report, or concepts that are in separate sentences. Therefore, we believe that by releasing these annotations, we will enable other researchers to make use of the data for many possible downstream applications, not only in the biomedical field, but also in application that rely on information extraction.

3. Annotations in EEG Reports

Only five types of medical concepts were annotated in the EEG reports: (1) EEG activities, (2) EEG events, (3) medical problems, (4) medical treatments, and (5) medical tests. Many of the medical concepts mentioned in the EEG reports are similar to those evaluated in the 2010 i2b2 challenge (Uzuner et al., 2011), as they represent medical problems, tests and treatments, thus we could take advantage of our

Concept Type	Polarity	Modality	EEG Activity-Specific
EEG Activity	✓	✓	✓
EEG Event	✓	✓	
Problem	✓	✓	
Test	✓	✓	
Treatment	✓	✓	

Table 1: Medical concept types and their attribute types.

participation in that challenge and use their definitions for identifying medical concepts. However, EEG reports also contain a substantial number of mentions of EEG activities and EEG events, as they discuss the EEG test. We defined these new types of medical concepts by relying on the definitions of the International Federation of Clinical Neurophysiology (Noachtar et al., 2004).

For each type of medical concept, we annotated their *modality* and *polarity*. We selected these two attributes because we have noticed that they are often expressed in the EEG reports. When we considered the recognition of the modality, we took advantage of the definitions used in the 2012 i2b2 challenge (Sun et al., 2013) on evaluating temporal relations in medical text. In that challenge, modality was used to capture whether a medical event discerned from a medical record actually happens, is merely proposed, mentioned as conditional, or described as possible. We extended this definition such that the possible modality values of “factual”, “possible”, and “proposed” indicate that medical concepts mentioned in the EEGs are actual findings, possible findings and findings that may be true at some point in the future, respectively. For identifying polarity of medical concepts in EEG reports, we relied on the same definition used in the 2012 i2b2 challenge, considering that each concept can have either a “positive” or a “negative” polarity, depending on any absent or present negation of its finding. Through the identification of modality and polarity of the medical concepts, we aimed to capture the neurologist’s beliefs about the medical concepts mentioned in the EEG report. For example, in the EEG report illustrated in Figure 1, we identified the medical problem *“right leg swelling"* with a “factual" modality and a “positive" polarity in the *clinical history* section. In the same section, the medical problem *“seizures"* had the modality “factual" and the polarity “negative".The medical problem *“hepatic encephalopathy"* in the *clinical correlation* section was found to have the modality “possible" and the polarity “positive".

3.1. Medical Concepts and their Attributes in EEG Reports

An EEG activity is defined as “an EEG wave or sequence of waves”, while an EEG event is defined as “a stimulus that activates the EEG” by the International Federation of Clinical Neurophysiology(Noachtar et al., 2004). EEG activities are describing the signals produced by the brain. The annotation of EEG activities in clinical narratives poses several challenges. As first reported in (Maldonado et al., 2017a), we noticed that EEG activities are not mentioned in a continuous expression. For example, in the narrative fragment: *“there are also bursts of irregular, frontally predominant [sharply contoured delta activity]$_{ACT}$, some of which seem to have an underlying [spike complex]$_{ACT}$ from the left mid-temporal region"* we can recognize one EEG activity that is mentioned by distant text spans in the narrative. To address this problem, we considered that EEG activities are expressed in EEG reports through (a) an *anchors* and (b) the attributes of the EEG activity. This assumption enabled us to identify the anchors of EEG activities, annotate them in EEG reports while also recognizing the attributes of EEG activities, without annotating them in the reports, but attaching them to the anchors. For this purpose, we defined 16 attributes which are specific to EEG activities.

Since the MORPHOLOGY best defines the EEG activities, we decided to use it as an anchor for each mention of an EEG activity in the EEG report. However, MORPHOLOGY represents the type or "form" of an EEG activity, which may have multiple values, as seen in Table 2, therefore the MORPHOLOGY remains also as an attribute of the EEG activities. When considering the MORPHOLOGY of EEG activities, we relied on a hierarchy of values, distinguishing first two types: (1) Rhythm and (2) Transient. In addition, the Transient type contains three sub-types: Single Wave, Complex and Pattern. Each of these sub-types can take multiple possible values, illustrated in Table 2. In addition to MORPHOLOGY, we considered three classes of attributes for EEG activities, namely (a) general attributes of the waves, e.g. the FREQUENCY BAND , the MAGNITUDE, and BACKGROUND - which asserts whether the EEG activity occurs in the background or not; (b) temporal attributes and (c) spatial attributes. The only temporal attribute considered is RECURRENCE, which describes how often the EEG activity occurs. As spacial attributes, we considered the DISPERSAL, the HEMISPHERE and eight additional attributes for the BRAIN LOCATION where the EEG activity is observed, since an activity can simultaneously occur in more than one brain location. All attributes have multiple possible values associated with them. Table 2 defines each of the 16 attributes of EEG activities and illustrates the possible values each of these attributes. Therefore, any identification of EEG activities in EEG reports amounts to recognizing the 16 attributes listed in Table 2 along with the polarity and modality attributes. We note that we selected these 16 attributes for EEG activities after consulting numerous manuals of neurology, and discussing with practicing neurologists. Furthermore, we were interested to generate a hierarchical structure for the Hierarchical epileptiform Activity Descriptors (HAD), rooted in the ACT tag. In deciding the nodes of the HAD, we have consulted the Epilepsy Syndrome and Seizure Ontology (ESSO)[2], which encodes 2,705 classes with an upper ontology targeting epilepsy and selected the concepts that best describe EEG activities.

In contrast, EEG events, which are frequently mentioned in EEG reports as well, can be recognized only by identifying the text span where they are mentioned and their polarity and modality attributes. While the Hierarchical Event Descriptors (HED) (available from http://www.hedtags.org) have defined many types of EEG experimental events, we decided to annotate the EEG events as a single tag Ev in the TUH corpus.

[2] http://bioportal.bioontology.org/ontologies/ESSO

<table>
<tr>
<td rowspan="6">*Attribute 1*: **Morphology** ::= represents the type or "form" of EEG
waves.
• Rhythm: continuous rhythmic activity
• Transient
  • Single Wave:
    • V wave
    • Wicket spikes
    • Spike
    • Sharp wave
    • Slow wave
  • Complex: A sequence of two or more waves having a characteristic form or recurring with a fairly consistent form, distinguished from background activity
    • K-complex
    • Sleep spindles
    • Spike-and-sharp-wave complex
    • Spike-and-slow-wave complex
    • Sharp-and-slow-wave complex
    • Triphasic wave
    • Polyspike complex
    • Polyspike-and-slow-wave-complex
  • Pattern: any characteristic EEG Activity
    • Suppression
    • Amplitude Gradient
    • Slowing
    • Breach Rhythm
    • Benign Epileptic Transients of Sleep (BETS)
    • Photic Driving (response)
    • Periodic Lateralized Epileptiform Discharges
    • Generalized Periodic Epileptiform Discharges
    • Epileptiform discharge (unspecified)
    • Disorganization
    • Positive Occipital Sharp Transients of Sleep (POSTS)
    • Unspecified: the default attribute value, used if no morphological information is given</td>
<td>*Attribute 2*: **Frequency Band**
• Alpha (8 – 13 Hz)
• Beta (13 – 32 Hz)
• Delta (< 4 Hz)
• Theta (4 – 8 Hz)
• Gamma (> 32 Hz)
• N/A</td>
</tr>
<tr><td>*Attribute 3*: **Background**
• Yes
• No</td></tr>
<tr><td>*Attribute 4*: **Magnitude** ::= describes the amplitude of the EEG activity if it is emphasized in the EEG report
• Low: e.g. subtle (spike); small (polyspike discharges)
• High: e.g. high amplitude (spike); excess (theta)
• Normal: the default value</td></tr>
<tr><td>*Attribute 5*: **Recurrence** ::= describes how often the EEG activity occurs
• Continuous: the activity repeats in a continuous, uninterrupted manner
• Repeated: the activity repeats intermittently
• None: the activity occurs once</td></tr>
<tr><td>*Attribute 6*: **Dispersal** ::= describes the spread of the activity over regions of the brain
• Localized (focal): limited to a small area of the brain
• Generalized (diffuse): occurring over a large area of the brain or both sides of the head
• N/A: the default value</td></tr>
<tr><td>*Attribute 7*: **Hemisphere** ::= describes which hemisphere of the brain the activity occurs in
• Right
• Left
• Both
• N/A</td></tr>
<tr><td colspan="2">*Location Attributes*: **Brain Location** ::= describes the region of the brain in which the EEG activity occurs. The BRAIN LOCATION attribute of the EEG Activity indicates the location/area of the activity (corresponding to the electrode placement under the standard 10-20 system).
• *Attribute 8*: Frontal (i.e. Anterior): The frontal region of the brain including all F*, Fp* and AF* electrodes
• *Attribute 9*: Occipital (i.e. Posterior): The occipital region of the brain including all O* electrodes
• *Attribute 10*: Temporal: The temporal region of the brain including all T* electrodes
• *Attribute 11*: Central: The central region of the brain including all C* electrodes
• *Attribute 12*: Parietal: The parietal region of the brain including all P* electrodes
• *Attribute 13*: Frontocentral: The area between the frontal and central regions of the brain including all FC* electrodes
• *Attribute 14*: Frontotemporal: The area between the frontal and temporal regions of the brain including all FT* electrodes
• *Attribute 15*: Centroparietal: The area between the central and parietal regions of the brain including all CP* electrodes
• *Attribute 16*: Parieto-occipital: The area between the parietal and occipital regions of the brain including all PO* electrodes</td></tr>
</table>

Table 2: Attributes specific to EEG activities.

EEG reports from the TUH corpus mention three other forms of medical concepts, which were used in the 2010 i2b2 challenge (Uzuner et al., 2011), namely *medical problems* (e.g., disease, injury), *tests* (e.g., diagnostic procedure, lab test), and *treatments* (e.g., drug, preventive procedure, medical device). These additional forms of medical concepts can be recognized by the text span where they are mentioned and their polarity/ modality attributes. The five medical concept types that we have annotated and their specific attributes are listed in Table 1 and in Table 2 details the EEG activity-specific attributes. In summary, we have annotated the medical concept types and their corresponding attributes which are summarized in Table 1. EEG activities, which capture the brain signals, are characterized by 18 different attributes: the 16 EEG-activity-specific attributes listed in Table 2 as well as polarity and modality. EEG events, medical problems, tests and treatments are characterized by only 2 attributes, namely polarity and modality.

3.2. Relations between Medical Concepts in EEG Reports

The relations that we have annotated in the TUH corpus of EEG reports are *binary relations* between the five types of medical concepts that we have considered, namely: (1) EEG activities; (2) EEG events; (3) medical problems; (4) tests; and (5) treatments. When selecting the types of re-

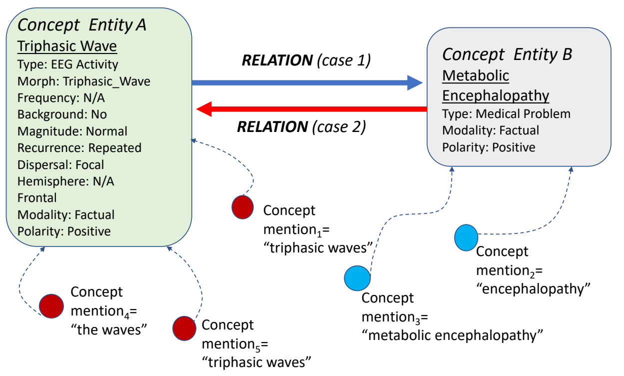

Figure 2: Concept entities, their mentions and possible relations between them.

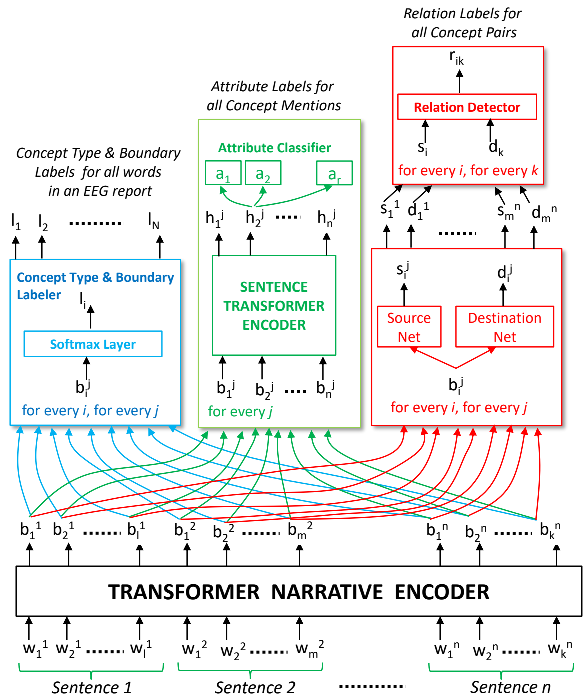

Figure 3: Architecture for Self-Attention Concept, Attribute and Relation (SACAR) Identification.

lations to be annotated, we have been informed by discussions with practicing neurologists and settled on four types of relations: (1) EVIDENCES; (2) EVOKES; (3) CLINICAL-CORRELATION and (4) TREATMENT-FOR. The motivation of selecting these relations was based on the observation that these relations correspond to the implicit knowledge gleaned by neurologists from the EEG reports which informs their reading of the Impression and Clinical Correlation sections of the EEG report. This relation schema is adapted from the schema reported in previous work (Maldonado et al., 2017b; Maldonado et al., 2018). The EVIDENCES relation considers EEG activities or medical problems as providing evidence for medical problems mentioned in the EEG report. The EVOKES relation represents the relationship where a medical concept evokes an EEG activity. EEG events, medical problems and treatments can all evoke EEG activities. The CLINICAL-CORRELATION relation connects the EEG activities and medical problems mentioned in the Clinical Correlation section of the EEG report if the activity *clinically correlates* with the medical problem. The TREATMENT-FOR relation links treatments to the medical problems for which they are prescribed.

When producing the annotations of relations, we took into account the observation that in the same EEG report, the medical concepts may be mentioned several times, which prompted us to distinguish between *concept mentions* and *concept entities*. Figure 2 illustrates two concept entities and their corresponding concept mentions. As it can be seen, the concept entities are illustrated through (1) their normalized name; (2) their type; and (3) their identified attributes. We normalized each concept mention into a canonical form (referred to as *normalized name*) using the (i) the morphology attribute for EEG activities and (ii) the United Medical Language System (UMLS) (Lindberg et al., 1993) *preferred name* of the concepts of other types. To identify the concept entities, we assumed that concept mentions from the same EEG report that (1) have the same normalized name, (2) the same type, and (3) the same values of their attributes *co-refer* to the same concept entity. We annotated relations between concept entities, not between concept mentions.

4. Joint Learning to Automatically Annotate Concepts, Attributes and Relations in EEG Reports Using Transformers

In (Maldonado and Harabagiu, 2019) we have detailed a neural network architecture capable of extracting (a) medical concepts; (b) their attributes and (c) the relations between them simultaneously from EEG reports. The architecture uses *self-attention* to learn a representation of all the words in the EEG report. Self-attention, sometimes called intra-attention, is an attention mechanism relating different positions of a single sequence (of words) in order to compute a representation of the sequence. The system reported in (Maldonado and Harabagiu, 2019) is called the Self-Attention Concept, Attribute and Relation (SACAR) identifier and it was used for automatically annotating concepts, their attributes, and relations spanning them by encoding long contexts across multiple sentences from the EEG reports available from the TUH corpus. SACAR employs a transformer narrative encoder (TNE) to generate an encoding, b_i^j, for each word, w_i^j, in each sentence $Sentence_j$ in the narrative of the EEG report[3]. These word encodings serve as input to: (1) *the concept type and boundary labeler* that detects the boundaries of each concept mentioned in an EEG report as well as its concept type; (2) *the attribute classifier* that recognizes the attributes of each medical concept; and (c) *the relation detector* that identifies relations between the concepts from the same EEG report. As shown in Figure 3, all these three modules operate jointly because they share the encoding of the words and sentences produces by the TNE.

[3] Tokenization was performed using the GENIA tagger (Tsuruoka et al., 2005) and sentence splitting was performed using OpenNLP (opennlp.apache.org).

Annotations	EEG Activity	EEG Event	Medical Problem	Test	Treatment
	Concept Types				
Gold	1438	452	798	716	539
Predicted	191,541	60,206	115,630	95,371	69,794
	Relation Types				
Annotations	EVIDENCES	EVOKES	TREATMENT-FOR	CLINICAL-CORRELATION	
Gold	397	342	356	195	
Predicted	49,942	45,554	47,419	25,974	

Table 3: Data statistics for the evaluation dataset along with automatically generated annotations.

Concept Type	Precision	Recall	F_1
EEG Activity	0.9460	0.9080	0.9266
EEG Event	0.9072	0.8299	0.8668
Medical Problem	0.8866	0.8358	0.8605
Treatment	0.9511	0.8814	0.9149
Test	0.9050	0.9280	0.9164
All Types (Macro Avg.)	0.9192	0.8766	0.8974

Table 4: Evaluation Results for Concept Type and Boundary Recognition.

Model	Precision	Recall	F_1
EVOKES	0.8402	0.8569	0.8485
EVIDENCES	0.5955	0.6109	0.6031
TREATMENT-FOR	0.5806	0.9042	0.7071
CLINICAL-CORRELATION	0.8287	0.8006	0.8144
All Relations (Macro Avg.)	0.7112	0.7932	0.7499

Table 5: Evaluation Results for Relation Identification.

The transformer narrative encoder (TNE) learns a contextualized encoding for each word in an EEG report given the full context of the EEG report using self-attention. In EEG reports, certain words tend to be more ambiguous than others, suggesting that computing the encodings of these words requires additional processing to correctly capture their meaning from the contexts in which they appear. Therefore the TNE leverages Adaptive Computation Time (Graves, 2016) to dynamically allocate more computational resources for the encoding of some words compared to others in the same EEG report. Specifically, the TNE consists of an Adaptive Universal Transformer (Dehghani et al., 2018) with 12 recurrent blocks as described in (Maldonado and Harabagiu, 2019). The TNE learns how many processing steps (blocks) are necessary to encode each word and to dynamically halt processing for one word, but to continue processing for other words whose encodings still require further refinement. The TNE is described in detail in (Maldonado and Harabagiu, 2019). The word embeddings produced by the TNE are shared among the three prediction modules.

The first prediction module, the Concept Type and Boundary Annotator, first identifies spans of text that correspond to medical concept mentions by assigning a label to each word in the narrative of the EEG report indicating that the word begins a medical concept mention (B), is inside of a medical concept mention but not at the beginning (I), or is outside of a medical concept mention (O). In this way, medical concept mentions can be identified by continuous sequences of words starting with a word labeled B optionally followed by words labeled I. In order to also identify the *type* of the medical concepts from the EEG reports and to distinguish EEG activities, EEG events, medical problems, treatments, and tests, we extended the IOB labeling system to include separate B and I labels for each concept type, yielding 11

Attribute	Precision	Recall	F_1
Morphology	–	–	0.8868
Rhythm	0.9549	0.9961	0.9751
V wave	0.7143	1.0000	0.8333
Spike	0.7500	0.3023	0.4309
Sharp wave	0.6522	0.8333	0.7317
Slow wave	0.9048	0.8261	0.8636
K-complex	1.0000	1.0000	1.0000
Sleep spindles	0.8333	1.0000	0.9091
Spike-and-slow	1.0000	0.6250	0.7692
Triphasic wave	1.0000	0.8000	0.8889
Polyspike complex	0.4000	1.0000	0.5714
Suppression	1.0000	0.2857	0.4444
Slowing	0.8710	0.9000	0.8852
Breach Rhythm	1.0000	1.0000	1.0000
Photic Driving	1.0000	1.0000	1.0000
PLEDs	0.3333	0.2500	0.2857
Epileptiform discharge	0.7000	0.6364	0.6667
Disorganization	0.7895	0.6250	0.6977
Unspecified	1.0000	0.4286	0.6000
Frequecy Band	–	–	0.9356
Alpha	0.9286	0.8387	0.8814
Beta	0.8750	0.6364	0.7368
Delta	0.9524	0.7143	0.8163
Theta	0.9000	0.6923	0.7826
N/A	0.9744	0.9954	0.9848
Background	0.9704	0.9428	0.9564
Magnitude	–	–	0.9273
Low	0.7273	0.4444	0.5517
High	0.8421	0.4848	0.6154
Normal	0.9636	0.9934	0.9783
Recurrence	–	–	0.9045
Continuous	0.6750	0.8438	0.7500
Repeated	0.7179	0.8740	0.7883
None	0.9761	0.9574	0.9667
Dispersal	–	–	0.8951
Localized	0.7714	0.5094	0.6136
Generalized	0.8400	0.4118	0.5526
N/A	0.9464	0.9943	0.9698
Hemisphere	–	–	0.9117
Right	0.8182	0.6429	0.7200
Left	0.7647	0.5200	0.6190
Both	0.8182	0.4737	0.6000
N/A	0.9540	0.9920	0.9727
Modality	–	–	0.9550
Factual	0.9682	0.9682	0.9682
Possible	0.7824	0.5027	0.6121
Proposed	0.7429	0.5098	0.6049
Polarity	0.9165	0.9071	0.9118
Location	0.7506	0.6062	0.6707

Table 6: Evaluation Results for Attribute Classification.

total medical concept boundary and type labels: L_C ={*B-ACT, I-ACT, B-EV, I-EV, B-PR, I-PR, B-TR, I-TR, B-TE, I-TE, O*}. The concept type and boundary labeler assigns a label $l_i \in L_C$ to each word w_i in the EEG report by passing each word's encoding produced by the TNE through a fully connected softmax layer as described in (Maldonado and Harabagiu, 2019).

As medical concepts are identified in EEG reports, the attribute classifier determines the values of each attribute type for each medical concept using another transformer encoder – this time operating at the sentence level – and a series of linear classifiers, one for each attribute type. The attribute classifier needs to further refine the encoding for each word using the *sentence transformer encoder* (STE), which is another adaptive universal transformer module similar to the TNE, operating at the sentence level instead of the full narrative. The word encodings produced by the TNE are fed to the STE, one sentence at a time. The encoding produced by the STE for the head token of each concept identified by the Concept Type and Boundary Annotator is then fed to a series of softmax classifiers to predict the attribute values for that concept.

The Relation Detector automatically identifies relations between pairs of concept entities recognized in EEG reports using a learned Bi-affine transform between mentions of each concept entity pair. More specifically, relation discovery was cast as a prediction of the most likely relation type between two concept entities. The two cases of selecting in each pair of concepts the relation source and destination, which are illustrated in Figure 3.2., were addressed simultaneously, considering that the first concept of the pair is the source while the second concept is the destination for one possible relation, and conversely for a second possible relation. The prediction of the most likely relation was made possible by generating for each concept mention c_i^k of every concept entity both a *source encoding*, s_i^k and a destination encoding d_i^k using the Source Net, and the Destination Net, respectively, as illustrated in Figure 3, implemented as two-layer neural networks:

$$s_i^j = \text{SourceNet}(c_i^j) = W_S^1 \left(\text{ReLU}\left(W_S^0 c_i^j\right)\right)$$
$$d_i^j = \text{DestinationNet}(c_i^j) = W_D^1 \left(\text{ReLU}\left(W_D^0 c_i^j\right)\right)$$

where $W_S^0, W_S^1, W_D^0, W_D^1 \in R^{d \times d}$ are weight matrices. These encodings enabled us to represent all possible relations between each pair of concept entities from an EEG report in terms of their mentions using an $N \times R \times N$ tensor, L, where N is the number of concept mentions discovered in the EEG report and R is the number of possible *relation types*. For each source concept entity S and each destination concept entity D, we consider (a) the source encodings of all the mentions of S in the EEG report, denoted as s_m; and (b) the destination encodings we produced for all the mentions of D, denoted as d_n. In order to compute each value of L, we used a bi-affine function between a source encoding, s_m, and a destination encoding, d_n, for each relation type $r \in R$. Details of the computation of bi-affine function, of the scores of all possible relations between each pair of concept entities as well as the way in which the probability distribution over all possible relation types are available in (Maldonado and Harabagiu, 2019).

5. Annotation Results

The SACAR model was used to automatically identify concepts, their attributes, and relations between them in the TUH EEG corpus. The TUH EEG corpus consists of 23002 documents. In this corpus, the average number of tokens per document is 264. Some EEG reports are very short, others are long, but the average number of sentences in EEG reports is 16. However, the longer EEG reports had on average 4,631 tokens, which about 17.5 times the size of average EEG report. The largest number of sentences in any EEG report was 230.

Annotation statistics for (a) the manually annotated data used to train SACAR and (b) the predicted annotations produced by SACAR are presented in Table 3. We evaluated the predicted annotations using a subset of 140 EEG reports from the TUH EEG corpus in which concepts, attributes, and relations were manually annotated by experts. This subset was constructed from a seed set of 40 manually annotated reports using ten rounds of active learning in which 10 reports were sampled each round as described in (Maldonado and Harabagiu, 2019). The evaluations were performed using 7-fold cross-validation on the gold subset of 140 EEG reports. Manual annotation was performed by three graduate students after extensive consultation with practicing neurologists. Average inter-annotator agreement, measured using Jaccard Score (Levandowsky and Winter, 1971), was 0.9658, 0.9518, and 0.8843 for concept boundary, attribute, and relation identification, respectively.

In our experiments we evaluated (1) the results of the concept annotation predictor; (2) the prediction of concept attributes both for EEG-specific concepts and generally clinically-relevant concepts; and (3) relation prediction. The results for concept type and boundary detection are presented in Table 4 in terms of precision, recall, and F_1 score, where predicted concept boundaries are considered correct if they exactly match a manually annotated boundary. Attaining a macro-averaged F_1 score of 0.8974, SACAR is able to accurately identify medical concept mentions in EEG reports.

Table 6 presents the results for each value of each attribute. Attribute values with no examples in the manually annotated (gold) data are omitted. Aggregated metrics are presented for each attribute type using micro-average, however precision and recall are omitted for multi-class classification tasks since they are equivalent to F_1. In general, performance for each attribute type is promising. However for some attributes, performance is significantly hindered (e.g. Morphology=Spike, Magnitude=Low, Dispersal=Generalized). This could be due to the scarcity of such attribute values in the annotated data. Future work may benefit from active learning policies which bias toward under-represented attribute values.

The results for relation identification are presented in Table 5. It is not surprising to find that the best predicted relations were the EVOKES and the CLINICAL-CORRELATION relations. In each abnormal EEG report, the EEG activities evoke some pathology, well known to practicing neurologists. Similarly, abnormal brain signals clinically correlate with known pathologies, thus such examples are abundant in the EEG reports. Conversely, EVIDENCES relations were the most difficult to predict, with SACAR attaining F_1 score of 0.6031. This is not surprising, given that EVIDENCES relations are often complex and even spark disagreement among neurologists (Beniczky et al., 2013). It is not uncommon for an activity to be noted as evidencing a medical problem in one EEG report, but not evidencing the same problem in another report. Moreover, the textual cues indicating some EVIDENCES relations varied and were often subtle. Furthermore, in many cases we noticed the need for accurate co-reference resolution of EEG activities mentioned in the Impression section and those mentioned in the Clinical Correlation.

6. Conclusion

We have described the annotation effort that captures the language used for describing brain signals and their correlations to other medical concepts. The corpus on which we have performed the annotations consists of a large set of EEG reports, available from the Temple University Hospital. We have annotated five types of concepts: EEG activities (representing the brain signals), EEG events, medical problems, treatments and tests. For all brain signals, we have also produced annotations for 18 different attributes, while for the other concepts we have annotated only modality and polarity attributes. We have also considered several types of relations between concepts in EEG reports, which were always long-distance relations. We have also shown that a self-attention-based neural model was successful at predicting not only annotations of EEG-specific concepts that describe brain signals and their attributes, but also annotations of clinically-relevant long-distance relations. These promising results should lead to the development of knowledge capture techniques operating both of Electronic Health Records and on medical literature, capable to characterise the language of brain signals.

7. Acknowledgements

Research reported in this publication was supported by the National Human Genome Research Institute (NHGRI) of the National Institutes of Health under award number U01HG008468, respectively. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

8. Bibliographic References

Beniczky, S., Hirsch, L., Kaplan, P., Pressler, R., Bauer, G., Aurlien, H., Brogger, J., and Trinka, E. (2013). Unified eeg terminology and criteria for nonconvulsive status epilepticus. 54 Suppl 6:28–9, 09.

Dehghani, M., Gouws, S., Vinyals, O., Uszkoreit, J., and Kaiser, Ł. (2018). Universal transformers. In 6th International Conference on Learning Representations, ICLR 2018, Conference Track Proceedings.

Goodwin, T. R. and Harabagiu, S. M. (2016). Multi-modal patient cohort identification from eeg report and signal data. In AMIA Annual Symposium Proceedings, volume 2016, page 1794. American Medical Informatics Association.

Graves, A. (2016). Adaptive computation time for recurrent neural networks. In arXiv preprint arXiv:1603.08983.

Levandowsky, M. and Winter, D. (1971). Distance between sets. *Nature*, 234(5323):34.

Lindberg, D. A., Humphreys, B. L., and McCray, A. T. (1993). The unified medical language system. *Yearbook of Medical Informatics*, 2(01):41–51.

Maldonado, R. and Harabagiu, S. M. (2019). Active deep learning for the identification of concepts and relations in electroencephalography reports. *Journal of biomedical informatics*, 98:103265.

Maldonado, R., Goodwin, T. R., and Harabagiu, S. M. (2017a). Active deep learning-based annotation of electroencephalography reports for cohort identification. In AMIA Summits on Translational Science Proceedings, volume 2017, pages 229–238. American Medical Informatics Association.

Maldonado, R., Goodwin, T. R., Skinner, M. A., and Harabagiu, S. M. (2017b). Deep learning meets biomedical ontologies: knowledge embeddings for epilepsy. In AMIA Annual Symposium Proceedings, volume 2017, page 1233. American Medical Informatics Association.

Maldonado, R., Goodwin, T., and Harabagiu, S. M. (2018). Memory-augmented active deep learning for identifying relations between distant medical concepts in electroencephalography reports. In AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science.

Noachtar, S., Binnie, C., Ebersole, J., Mauguière, F., Sakamoto, A., and Westmoreland, B. (2004). A glossary of terms most commonly used by clinical electroencephalographers and proposal for the report form for the eeg findings. *Klinische Neurophysiologie*, 35(1):5–21.

Smith, S. J. M. (2005). Eeg in the diagnosis, classification, and management of patients with epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry*, 76(suppl 2):ii2–ii7.

Sun, W., Rumshisky, A., and Uzuner, O. (2013). Evaluating temporal relations in clinical text: 2012 i2b2 challenge. *Journal of the American Medical Informatics Association*, 20(5):806–813.

Tsuruoka, Y., Tateishi, Y., Kim, J.-D., Ohta, T., McNaught, J., Ananiadou, S., and Tsujii, J. (2005). Developing a robust part-of-speech tagger for biomedical text. In Panhellenic Conference on Informatics, pages 382–392. Springer.

Uzuner, Ö., South, B. R., Shen, S., and DuVall, S. L. (2011). 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association*, 18(5):552–556.