2022 IEEE International Conference on Big Data (Big Data) | 978-1-6654-8045-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/BigData55660.2022.10020994

Mining Latent Disease Factors from Medical Literature using Causality

Pranav Gujarathi[1], Jack T. VanSchaik[2], Venkata Mani Babu Karri[2], Anushri Rajapuri[2], Biju Cheriyan[2]
Thankam P. Thyvalikakath[2], Sunandan Chakraborty[2]
[1]*Indiana University Bloomington*
[2]*Indiana University IUPUI*
Indianapolis, IN
{pgujarat,jtvansch,vkarri,anraja,bijcheri,tpt,sunchak}@iu.edu

***Abstract*—Understanding causality is a longstanding goal across many different domains. Different articles, such as those published in medical journals, publish newly discovered knowledge, often causal. In this paper, we use this intuition to build a model that leverages causal relations to unearth factors related to Sjögren's syndrome. Sjögren's syndrome is an autoimmune disease affecting up to 3.1 million Americans. The uncommon nature of the disease, coupled with common symptoms of other autoimmune conditions such as rheumatoid arthritis, it is difficult for clinicians to timely diagnose the disease. This is further worsened by suboptimal communication between dentists, and physicians, including rheumatologists and ophthalmologists, because clinical manifestations of this disease require the patients to visit physicians with different specialties. A centralized information system with easy access to common and uncommon factors related to Sjögren's syndrome may alleviate the problem. We use automatically extracted causal relationships from text related to Sjögren's syndrome collected from the medical literature to identify a set of factors, such as "signs and symptoms" and "associated conditions", related to this disease. We show that our approach is capable of retrieving such factors with high precision and recall values. Comparative experiments show that this approach leads to 25% improvement in retrieval F1-score compared to several state-of-the-art biomedical models, including BioBERT and Gram-CNN.**

***Index Terms*—causal relationships, relationship extraction, text mining, sjogren's syndrome**

I. Introduction

Causal relationships depict essential knowledge across many different fields, including medicine and health. Researchers in these fields design and conduct experiments to establish causality between two events and publish their findings in research articles. Such research articles record the discovery of new causal relationships or new conditions for existing relationships. Thus, mining such relationships from authentic text, such as published research articles, provides a unique opportunity to aggregate causal knowledge in a field. This is particularly true in the area of health and medicine, where for many diseases, diagnosticians are not aware of *all* factors associated with a disease; as a result, diagnosis is often delayed. An example of such a disease is Sjögren's syndrome. Sjögren's syndrome is an autoimmune disorder where the immune system destroys glands that produce tears and saliva [1, 2] and is also associated with rheumatic disorders [3, 4, 5]. Most people with Sjögren's syndrome have limited symptoms, such as dry eyes and mouth. Their general but lack of timely intervention may affect other body organs, including the kidneys, blood vessels, lungs, liver, pancreas, and brain [6]. Moreover, the primary symptoms for Sjögren's syndrome are spread across several domain areas, such as dentistry, ophthalmology, and rheumatology. This lack of continuity and suboptimal communication between dentists and physicians create a critical gap in adequately understanding the disease's characteristics. Hence, it becomes a challenge for clinicians to diagnose Sjögren's syndrome timely.

Several studies have addressed many research questions related to Sjögren's syndrome and published the findings in peer-reviewed journals. These journal articles contain significant results about Sjögren's syndrome concerning new symptoms, risks, and associated conditions [2, 7, 8, 9]. However, many such findings remain within the confines of those articles and are seldom used in practice [10]. The volume of these articles makes it hard to find little-known factors and use them effectively in diagnosis. In this paper, we present a novel information extraction method to retrieve such factors related to Sjögren's syndrome that will help clinicians across specialties to timely diagnose and treat patients with Sjögren's syndrome. The factors related to Sjögren's syndrome that may benefit clinicians can be classified into four categories – (1) signs and symptoms, (2) risk factors, (3) associated conditions, and (4) diagnostic tests [3, 11]. Among these four categories, relationships between "signs and symptoms" and "associate conditions" are often expressed using causal semantics. For example, *"Sjogren's syndrome can cause not only corneal perforation but also mucosal perforation which may lead to a lacrimal fistula"* [7] - this sentence expresses the possibility of two symptoms that might be *caused* by Sjögren's syndrome. In this paper, we present a novel method to identify causal sentences from research articles and use them to unearth little-known factors related to Sjögren's syndrome.

There are numerous way causality can be expressed in natural language text; as a result, extracting causal knowledge from text is a challenge. Causality can be stated explicitly (e.g., mosquito bite <u>causes</u> malaria) where the relationship is explicitly stated with a clear marker – *causes* [12, 13], as well as implicitly (e.g., Last week temperature rose significantly, there were several cases of heat stroke reported), without using causal markers. Past works have used many machines and deep learning based approaches [13, 14, 15, 16, 17], but they only target explicit causality and cannot holistically extract causal relationships. They also ignore that text presents causality through multi-word expressions or phrases instead of single words. Another drawback of these methods is that they essentially approximate a function, and we need more versatility to extract causal relationships from a vast text-domain. We propose an unsupervised framework for the causality extraction from sentences using Deep Q Reinforcement Learning (RL) method [18]. Given a sentence or a document, we aim to extract two sets of words and phrases connected by a causal semantic (which may not be explicit). To extract such words, we propose an RL agent that will iterate over multiple episodes (subsamples of data) and increase the chance of identifying the correct cause words or causal phrases along with the related effect words or phrases by maximizing a reward. We train and test the model on two separate datasets containing causal sentences, SemEval-2010 and ADE[19, 20] and apply the trained model on a different dataset built for Sjögren's syndrome from abstracts collected from the PubMed database. The causal relationship extraction model was performed with an F1-score of 0.89 and 0.87 on the SemEval-2010 and ADE datasets, respectively. We compared these results with several baseline models and related works that used the same dataset. We found that our model's performance was slightly lower than just one model [21] (F1-score: 90.6) on SemEval-2010 but outperformed the state-of-the-art models trained on the ADE dataset. We observe similar patterns while extracting factors related to Sjögren's syndrome. The precision and recall for our method in extracting Sjögren's syndrome-related factors were 0.85 and 0.78 respectively (F1-score: 0.81), which was at least 25% better than other state-of-the-art models, such as, Gram-CNN and BioBERT.

II. Background and Related Work

Researchers in many fields, design and conduct experiments using methods like, observational studies and randomized control trials to determine whether two events are causally linked, and scholarly articles publish newly discovered causal knowledge emerging from those studies. We see a broad spectrum of work that attempts to retrieve such known causal relationships from a large corpus of documents and apply them to problems like question answering [22], medical education [23], and financial analytics [24], among others. Expressing causality in a sentence may take several forms; the majority of them are *marked* but maybe *explicit* or *implicit*. Explicit causality has relations that are connected by: (a) causal links (e.g., hence, therefore); (b) causative verbs (e.g., causes, leads to); (c) conditional (e.g., if...then...) [25]. The sentence: "mosquito bites cause malaria," where the cause and effect are directly linked by the word "cause" [12, 13] is an example of explicit causality. Implicit causality involves using ambiguous connectives, e.g., *as*, *after* etc., as they are equally likely to be used in causal or non-causal context. For example, "as" is used as a causal marker in the sentence: "There was no debate <u>as</u> the Senate passed the bill on to the House" [12]. Some causal sentences may not have any connectives, for example, the sentence: "Last week temperature rose significantly, there were several cases of heat stroke reported"), where the relationship *rising temperature is the cause of the heatstroke cases* has no causal marker. These are called *unmarked* causal sentences. Causal relationships may span across the sentence. For example, the following two sentences depict a causal relationship [*financial stress* $\rightarrow$ *divorce*]: "Being unfaithful can lead to divorce. On the other hand, financial stress is another significant factor." [26]. Most approaches [16, 27, 28, 29] identify causality in the basic levels, which are explicit and/or intra-sentential forms.

Past works that addressed this problem can be broadly divided into three groups: rule-based, statistical machine learning(ML)-based, and deep learning-based approaches. Earlier works were mainly rule-based, where linguistic patterns were used to detect explicit causality [28, 30]. Girju et al. [31] devised a novel approach to a rule-based system, where linguistic patterns were automatically learned instead of manually setting up the rule base. Inspired by previous works that used lexico-syntactic patterns to infer causation, a new suite of ML-based methods emerged. Meuller et al. [32] presented a novel method and a working prototype that automatically extracts both causes and effects and signs, mediators, and conditions from scientific papers. CausalTriad [33] used a minimally supervised approach, using distributional similarity and discourse connectives. Few other works exploited linguistic structures, such as multi-word expressions [34], N-grams, topics and sentiments [35], lexical patterns [31, 36].

With the emergence of deep learning methods, we observe their application in extracting causal relationships from the text [14, 37, 38]. Xu et al. [39] used LSTM to learn higher-level semantic and syntactic representations along the shortest dependency path, while Li et al. [40] combined BiLSTM with multi-head self-attention to direct attention to long-range dependencies between words. The latter showed significant improvement when the cause-effect words have greater separation. Some studies demonstrate that attention, especially of the multi-attention mechanism, shows better performance [40, 41]. In addition to RNN, we observe the use of CNN; an example is by Wang et al. [41], who proposed a multi-level attention-based CNN model to capture entity-specific and relation-specific information and the use of graph-based deep learning models, such as GCN. Zhang et al. [42] proposed a dependency tree-based GCN model to extract relationships. Recently, we have seen the application of contextual word embeddings and large pre-trained language models in this space. Kyriakakis et al. [16] used BERT [43] and ELMO [44] showed that with large datasets, these models could improve previous state-of-the-art performance. Zhang et al. [45], combined LSTM with entity position-aware attention to encode both semantic information and global positions of the entities.

III. Sjogren's Syndrome

Sjögren's syndrome (SS) is the second most common autoimmune connective tissue disease [1] affecting up to 3.1 million Americans [2]. It causes lymphocytic infiltration of salivary and lacrimal glands resulting in dry mouth and eyes. SS is common among middle-aged people, with a high prevalence in females (female: male 9:1) [46] [47]. The exact etiology of SS is not known [8] and is considered to be multifactorial due to endocrine, genetic and viral factors [48]. As a result, the diagnosis of SS is delayed, and this causes undue disease burden on Sjögren's syndrome patients (SSP) [49]. Despite this disease burden, SS has not attracted the same level of research rigor as other rheumatic diseases, such as rheumatoid arthritis and systemic lupus erythematosus (SLE) [49]. The unusual nature of the disease makes it difficult to conduct sufficiently powered prospective clinical studies longitudinally and determine the disease changes over time. A lack of continuity and communication between dentists, and physicians, including rheumatologists and ophthalmologists, is also a crucial reason for the poor understanding of SSPs' disease characteristics.

Moreover, SS shares common clinical manifestations with other autoimmune conditions, such as rheumatoid arthritis [50], and often SS is mistakenly diagnosed with these conditions, thus further delays the diagnosis [51]. Therefore, more research is needed to understand the SS disease progression and characteristics that identify clinical features distinguishing SS from other conditions leading to early diagnosis of SS and treatment. This can be facilitated by identifying certain factors associated with SS; this includes additional conditions that frequently co-occur with SS, risk factors, and diagnostic tests reported in the existing literature. Such a comprehensive list of information will determine the prevalence of these findings in the electronic health record (EHR) data of SSPs diagnosed and managed in real-world clinical settings. Thus, identifying uncommon or little-known factors extracted from the scientific literature will help to build a comprehensive a knowledge base of conditions that will help clinicians with more accurate and timely diagnosis of SS is harder than solely referring to clinical data.

This paper aims to establish a novel entity extraction model that automates the retrieval of clinical findings relevant to SS from the scientific literature. Such a model will support mining relevant information from a large corpus of literature, which is infeasible through a manual process. The factors related to SS that we aim to extract from the scientific literature, such as symptoms, associated conditions, and risks, are expressed in the literature using some form of *causal* semantics. For example, "dry mouth *is caused by* Sjogren's syndrome." Thus, our goal is to design a causal sentence classifier that identifies sentences with causal semantics and extracts the *cause* and *effect* event pairs from those sentences.

IV. Problem Statement

We define the problem of identifying causal relationships from natural language text as a sequence labeling task. If an input sentence with n words is represented as $X = x_1, x_2, ..., x_n$, then produce an output sequence of length n, $Y = y_1, ..., y_n$, where $y_i \in \{CAUS, REL, EFF, NONE\}$. The possible values of y represent the tags a word can be labeled with. Here, *CAUS* represents a word or phrase of words representing a causing event or factor, similarly *EFF* represents the effect of the causing factor, *REL* represents the relationship or the connective words/phrases that depict the causal relationship. As we do not assume marked and explicit causality, we will omit the *REL* tag. Finally, *NONE* represents parts of the input sentence that do not contribute to causality. Figure 1 shows an example causal sentence with different labels. In this example, the words 'Renal', 'tubular,' and 'acidosis' will have the label *CAUS*, and the label of 'hypokalemia' will be *EFF*. *NONE* will be assigned to the words, such as 'which,' 'can,' and 'be.' We assume that majority of the sentences from a document will not depict any causality and will have sequences of *NONE* as output. The conditional probability of our model can be depicted as,

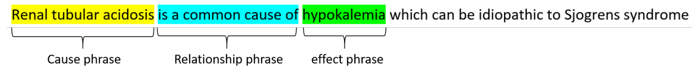

Fig. 1. An example sentence [2] with a causal relationship that highlights a factor that may lead to Sjögren's syndrome

$$p(Y|X) = \prod_{i=1}^{n} p(y_i|x_{1...i}, y_{i=1...i-1})$$

Addressing this problem will require a model that embodies the semantics of expressing causality while exploiting the syntax that allows that structure. Given the nature of this problem, a sequential model is more likely to perform better; however, past works have shown that LSTMs and similar models cannot exploit the syntactic structure of the input sequences [52]. While a purely semantics-based model can detect marked and explicit causal statements, identifying unmarked or implicit causality in the text is beyond its scopes. A sequential model assumes the start of the sequence from the beginning of a sentence. This is natural as humans perceive text in the same way, but information flow may follow different directions. The root information of a sentence is in the verb, which, according to the subject-verb-object structure followed in English is usually in the middle of the sentence. This middle verb section of the sentence will likely determine whether a statement is causal. We exploit this non-linear syntactical structure of a sentence by designing a reinforcement learning agent that traverses the sentence along these non-linear paths through adaptive *actions* and collecting *rewards* for the correct identification of labels.

To identify factors related to Sjögren's syndrome, we analyze sentences where $\mathcal{L}$("Sjögren's syndrome") $\in$ {*CAUS*,*EFF*}, where $\mathcal{L}(w)$ represents the label of the word w. If $\mathcal{L}$("Sjögren's syndrome") $=$ *CAUS* then $\forall w, \mathcal{L}(w) =$ *EFF*

[2]https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3822229/

will represent factors caused by Sjögren's syndrome and vice versa. To generalize this approach, we relax the above constraint and state, $\mathcal{N}($ "Sjögren's syndrome"$) \in \{CAUS, EFF\}$, where $\mathcal{N}(x)$ represents the noun phrase containing x. This relaxation will not fail to associate factors of Sjögren's syndrome found in sentences like in Fig 1, where "Sjögren's syndrome" is not directly part of *EFF*.

V. Causality Extraction using Reinforcement Learning

With new emerging Transformer models and architectures, as well as the availability of pre-trained models, fitting a model on annotated data is the easiest part of the process but impractical considering such annotated data is scarce and hardly available. Hence exploring unsupervised methods has an additional value - they open an avenue for a discussion where neatly annotated data is not a necessary blocker to getting started with creating a valuable natural language model. Specific domains often require subject experts to annotate properly and cannot be outsourced to popular annotation services. This is particularly true in our case, as labeling requires prior knowledge on Sjögren's syndrome.

Deep Reinforcement Learning in recent times has emerged as a promising approach that can utilize popular architectures (e.g., Transformers, CNNs, LSTMs, etc) while also going a step further than function approximation toward general Artificial Intelligence. This is possible because RL tasks are formulated as an optimization strategy. We simulate an agent playing a finite sequential game to improve the reward obtained at each step gradually. The critical difference is that this scalar reward neither needs ground truth labels nor has to be differentiable - as long as the reward magnitudes reflect the agent behaving favorably.

We propose an unsupervised framework for the causality extraction from sentences using A2C [53, 54] or Actor Advantage Critic Method. The advantage of this framework is that even though we lack ground-truth labels, essential for supervised learning, we can creatively use pre-trained models to assign a 'score' or evaluate our predictions. We use a combination of pre-trained models and hypotheses to formulate the score (μ_t). The basic idea behind this is that provided the predictions are correct, certain conditions must hold. For instance, if the predicted cause and effect are correct and they are indeed the phrases representing causal entities, firstly, these entities should be noun phrases or compound noun-like entities. Secondly, if we frame a new sentence using these entities, the 'conclusion' phrase should be consistent or 'agree' with the premise phrase.

A. Reinforcement Learning Steps

A typical RL problem consists of the following setup: a sequential task, where an *agent* starts at a initial position(s_0), and has to navigate through different *steps* to eventually reach an endpoint(s_T), which is referred to as completing an *episode*. At every step, the *agent* receives feedback on the decisions. Based on the feedback, at time t, it tries to take action (a_t) that will maximize the reward(r_t). Eventually, after multiple simulations of an *episode* and using an optimization algorithm, the objective is to maximize the cumulative *reward* or $\sum_{t=0}^{T} r_t$ for an *episode*. We use this setup and define s_t, r_t and a_t to ensure that maximizing $\sum_{t=0}^{T} r_t$ will improve the prediction accuracy of labeling words in a sentence as *cause* and *effect*. Table I summarizes the notations used in this model.

TABLE I
NOTATIONS USED IN THE REINFORCEMENT LEARNING SETUP

Notation	Description
t	Time step 't'
T	Maximum time steps in an episode.
υ	a random sub-sample of input sentences
θ_t	Given υ, represents predictions at time t
s_t	State at time step t(subject to definition)
a_t	Action taken at time step t. Action may not necessarily be the same as predictions, although they directly lead us to the prediction. For instance, action can be like softmax scores or probabilities, while θ_t are the actual textual prediction inferred from it.
μ_t	Given an υ and corresponding θ, this represents the score of the prediction or how good it is.
r_t	Represents reward at time step

B. RL task setup

For a particular episode, we pick a random subsample(υ) of sentences. At every step, the *agent*(in our case a neural network), takes s_t as an input and predicts a_t, also giving us θ_t (Table I). We score this prediction and assign value μ_t to it accordingly; since we want to use previous feedback and results to guide current action, we define the *state*(s_t) as a collection of a time-invariant variable (input sentence) as well as two time-dependent variables (previous state and scores) incorporating the information of the trajectory after the start.

$$s_t = [\upsilon, .a_{t-1}, \mu_{t-1}]$$

Since RL algorithms optimize $\sum_{t=0}^{T} r_t$, we define our reward as:

$$r_t = \mu_t - \mu_{t-1}$$

Thus, $\sum_{t=0}^{T} r_t$ is $\mu_t - \mu_0$, meaning optimizing cumulative reward is the same as improving the prediction score compared to a random walk (based on our definition). As mentioned earlier, we can leverage RL algorithms for unsupervised learning since there are ways to use pre-trained models creatively in a way that allows us to assign a score to a cause-effect prediction automatically.

We assume that if one of the phrases, cause or effect is known, we can frame questions to infer the other. Thus, for every cause phrase, we define question templates: if $\{0\}$ and $\{1\}$ represent the cause and effect phrases respectively, then we can frame questions like. [What causes $\{1\}$?, What does $\{0\}$ cause?, Does $\{0\}$ cause $\{1\}$?], say ω questions per every input sentence. We know that both cause and effect belong to the set of noun phrases and that if they occupy the places by $\{0\}$ and $\{1\}$ respectively, the questions can be answered with high confidence. We use the pre-trained QA model, which takes in two inputs - question and context and returns two inputs - answer phrase and confidence score. Hence for every sentence, we feed in ω different sets of inputs (same context, different question), and finally pick the one with the best confidence score. For example, for the sentence - "*the subjects were exposed to UV irradiation causing a local tissue inflammation*", the question *What is causing local tissue inflammation?* gave the best confidence score with the answer *uv radiation*. Hence for the assignment: *uv radiation* and *local tissue inflammation* respectively as cause and effect, the reward (r_t) will be higher. Figure 2 explains this process.

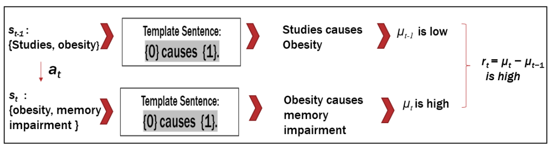

Fig. 2. Scoring method explained for the sentence "*Studies have supported that obesity accelerates AD-related memory impairement.*" using question templates

C. Actor Advantage Critique Algorithm (A2C)

The actor critique algorithm is based on Deep Q-learning Network (DQN) [18] algorithm. This RL framework is used along with actions and rewards designed based on our NLP tasks to extract cause and effect pairs. This network uses the Value function and Q-values at each state to compute the usefulness and quality of the state. At each state s_t consisting of υ , a_{t-1} and μ_{t-1} where sentence stays constant whereas a_{t-1} and μ_{t-1} are the feedback terms. The μ_{t-1} is a scalar output and a_{t-1} is a vector of $4\times\ maxlen(\upsilon)$. We fix that the maximum length of a sentence is 80 words for our experimentation, and we estimate the probability of every word to be a cause or an effect word. The output vectors for each word will have a size of four. Each element will represent the probability of the word being the start of a cause phrase $\phi^s(\kappa)$, the probability of the word being the end of a cause phrase $\phi^e(\kappa)$, probability of the word to be the start of an effect phrase $\phi^s(\epsilon)$, and probability of the word to be the end of an effect phrase $\phi^e(\epsilon)$ respectively. Based on this probability distribution, start and end indices of cause and effect phrases are determined.

D. Architecture and setup

The goal of our model is to identify cause and effect phrases. At each iteration of the state a sentence of length $len(d_i)$ is passed through Albert [55], a lighter version of BERT [43] based transformer model with 12 million parameters to generate sentence embeddings of size ($len(d_i)$,768). This output is then batch normalized [56] and is reduced by taking a mean across the length l resulting in a vector of size (1,768). Then the action a_{t-1} output from the previous state of size (80, 4) is reduced to (1,128) and batch normalized. This output $a^{|}_{t-1}$ and μ_{t-1} are combined into one single vector of size (1, 896), this output is further reduced and normalized to (1, 128) and combined with the scalar epsilon from previous state ϵ_{t-1}.

VI. Evaluation

We evaluate this work in two phases - (Task 1) evaluate the performance of the causal relationship extraction model, and (Task 2) validate the findings after applying this model to extract factors for Sjögren's syndrome.

A. Datasets

We use three different datasets to validate our approach. We use the first two datasets – SemEval-2010 Task 8 and Adverse Drug Effects (ADE) – to train and evaluate the causal relationship extraction model (Task 1) and a custom dataset built from articles related to Sjögren's syndrome collected from the PubMed database.

- **SemEval-2010 Task 8 (SE2010) [19]:** This dataset contains sentences depicting a set of seven semantic relationships, including cause-effect relations. It has in total of 1,331 sentences that relate to causal relationships.
- **Adverse Drug Effect (ADE) [20, 57]:** This dataset contains sentences explaining the adverse effects of drugs using causal sentences. It has been curated from 1,644 PubMed abstracts and contains 6,821 causal sentences. However, this dataset has minimal variation in terms of syntax and vocabulary, and in all sentences, the causality is expressed through the verb *"causes"* and its variation.
- **Sjögren's syndrome dataset (SSD):** We created this dataset from 2,350 PubMed abstracts retrieved using the keyword "Sjögren's syndrome" and its variants. This dataset has 26,525 sentences, and we selected a subset of 1,058 sentences for our analysis. The words in these 1,058 sentences were annotated using four labels – signs and symptoms, risk factors, associated conditions, and diagnostic tests. These sentences were simultaneously labeled by two annotators, both domain experts, and had prior experience working with Sjögren's syndrome. The annotators had 90.7% agreement in labeling these sentences. After manually inspecting these sentences, we found 383 sentences with causal semantics and used them for task 2.

B. Performance of the Causal Relationship Extraction Model

We evaluated our model on the SemEval and ADE datasets by comparing our findings with the ground truth labels. We use the metrics precision, recall, and F1-score scores to estimate the performance of identifying cause and effect from the sentences. In both the datasets, only the cause and effect are labeled and not relationship words; hence we ignore our model's *RES* labels during validation of the results. We compare our findings with several baseline models, such as Long Term Short Term neural memory network (LSTM) [58], Bidirectional LSTM (BiLSTM) [59]. These architectures allow the sequential information to flow in one direction (left to right) for LSTM and in both directions for BiLSTM. We pre-processed the sentences to word representations using pre-trained word embeddings from Glove vectors [60]. We add another baseline model using the BERT language model and the corresponding embedding [43]. We use a fully connected network as the final classifier to output the labels. Table II presents the performance across all these models. Our approach outperformed all these baseline models, and the F1-score is almost 6% better than the following best (BERT-based) model. Given that our approach has a minimal dependency on annotated data, the net performance boost is even more.

Many previous works have used the same datasets and developed causal relationships and extraction models. These works have used a combination of statistical machine learning and deep learning methods to identify causal relationships from the text. We identified the best-performing models from the literature for each dataset (SemEval and ADE) and compare our performance. Among the best performing model on SemEval-2010 is a variant of BiLSTM proposed by Li et al. [17]. They combined BiLSTM with multi-head self-attention to direct attention to long-range dependencies between words. Wang et al. [41] also used an attention-based model on CNN instead of BiLSTM. Presently, the best performing model trained on SemEval-2010 is by Kyriakakis et al. [21]. They used pre-trained language models, such as BERT [43] and ELMO [61] and used Bidirectional GRU with self-ATTention (BIGRUATT) as the base model. Experimental results show that BERT model combined with BIGRUATT performs better on most occasions and scales well with a larger dataset, where the base model reaches a plateau. Among the best-performing models trained on the ADE, the corpus includes the model proposed by Wang and Lu [62], where they focused on jointly modeling entities and relationships. They used a sequence and a table encoder to help each other jointly learn the entities and relations. Zhao et al. [63] used a similar joint modeling technique but proposed Cross-Modal Attention Network (CMAN) has two attention units consisting of BiLSTM-enhanced self-attention (BSA) and BiLSTM-enhanced label-attention (BLA) units. Table III presents the summary of this comparative analysis and shows the F1 scores in comparison to our approach. Our model outperformed other top models trained on ADE. On the other hand, for SemEval-2010, our model was marginally poorer than Kyriakakis et al. [21]. Considering the performance across datasets, our model is likely to perform at par or better than other models.

C. Identification of Factors related to Sjögren's syndrome

We apply the causal relationship extraction model tested on SemEval-2010 and ADE datasets on the Sjögren's Syndrome Dataset (SSD) to identify causal sentences and the corresponding cause and effect phrases to extract factors related to Sjögren's syndrome. The sentences in the SSD dataset were manually annotated with four labels. Out of 1,058 annotated sentences, we found 383 causal sentences, which are the potential candidates to unearth factors related to Sjögren's syndrome. As stated earlier, we assume causality can only identify a subset of the factors, and extracting the labels "risk factors" and "diagnostic test" is beyond its scope. Hence, we omitted these labels during the evaluation. We extract these factors by collecting the opposite label (either *cause* or *effect*) when the term "Sjögren's syndrome" or its variants is detected as *cause* or *effect*. We present a set of selected causal-effect pairs extracted through our model in Table IV. In these examples, we see that Sjögren's syndrome can appear as a cause as well as an effect, which represents the possibility of how factors associated with Sjögren's syndrome are mentioned in the text, and the capability of our method to detect them. In these selected examples, we see different types of factors (or labels), such as signs and symptoms (e.g., "loss of secretion," "xerophthalmia") and associated conditions (e.g. "annular erythema," "non-Hodgkin's lymphoma"). This This conforms to our assumption that causality is more likely to be used to represent the labels, *signs and symptoms* and *associated conditions*.

To verify the above claim, we applied our model to the labeled dataset where 1,058 sentences were annotated using four labels. We collected the *cause* (or effect) associated with the term "Sjögren's syndrome" when it is the *effect* (or cause) and computed the retrieval accuracy of those two labels. We created a test set containing 100 sentences out of the 383 causal sentences (the remaining sentences were used to train or fine-tune the baseline models). We retained only two labels, "signs and symptoms" and "associated conditions," for this experiment. This approach of using causal relationships to extract factors related to Sjögren's syndrome performed with a precision of 0.85 and recall of 0.78 (Table V).

We compared our findings with several baseline models designed for sequence labeling. These were supervised models, trained on a set of 283 annotated sentences, and tested on the same test set. Baseline models included BiLSTM and a modified BiLSTM, where the output layer of the BiLSTM model was fed into a CRF model. This approach has been shown to improve performance in other applications [17]. In addition, we have used several other models that have improved performance when trained on biomedical text, including BioBERT [64]. Experiments show that BioBERT has exceeded significantly in entity labeling and relationship extraction on biomedical text compared to many state-of-the-art models, including BERT [43]. We also applied BioWordVec [65], a biomedical language model built using fastText [66] and gram-CNN [67], which is trained for named entity recognition (NER) task for biomedical literature. The results from these experiments are summarized in Table V.

We validate our findings with the manually annotated ground truth labels. We use precision, recall, and F1 score to compare the performance of our approach. Table V summarizes the findings. Our approach is unsupervised; hence, it can identify the factors but cannot put a label (such as signs and symptoms) next to them. Moreover, we targeted only two out of four labels in the annotated set. We modified the supervised training for the baseline models and removed the two labels for

TABLE II
COMPARISON OF OUR REINFORCEMENT LEARNING METHOD WITH BASELINE MODELS

	SemEval-2010			ADE		
	Precision	Recall	F1 score	Precision	Recall	F1 score
Our approach (Reinforcement Learning)	**0.93**	**0.86**	**0.89**	**0.88**	**0.85**	**0.86**
LSTM-Glove	0.78	0.82	0.79	0.80	0.77	0.78
BiLSITM-Glove	0.82	0.80	0.81	0.82	0.84	0.82
BERT	0.87	0.83	0.84	0.86	0.831	0.84

TABLE III
COMPARISON WITH SELECTED RELATED WORKS

Dataset	Model	F1 Score
4SemEval-2010	Li et al. [17]	84.6
	Wang et al. [41]	88.0
	Kyriakakis et al. [21]	**90.6**
	Our approach	89.4
4ADE	Gurulingappa et al. [20]	70.0
	Wang and Lu [62]	80.1
	Zhao et al. [63]	81.1
	Our approach	**86.4**

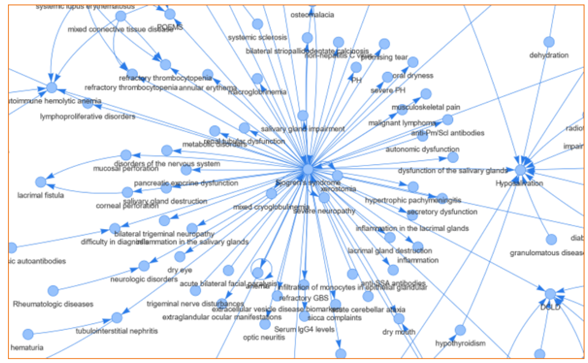

Fig. 3. Screenshot of a part of the causal network

a fair comparison. In addition, we compressed the confusion matrix for the baseline models and ignored any inter-label misclassifications. This means if a model misclassified a phrase to be “signs and symptoms” instead of “associated risk” and vice versa, we considered it a correct classification. This maintains parity with our approach, and the outcome is consistent across all models, where we output factors without a label (or name) associated with them.

We created a causal network by combining the individual cause-effect pairs. In this network, each cause-effect pairs were represented as two nodes connected by a directed edge from cause to effect. Then the nodes were merged based on similarity (i.e., same names) to have connected components combining the preliminary isolated pairs. This network provides additional information, such as chains of transitive causal relationships, and mediators, confounders through triangular structures. Through this causal network, we have observed that *Tubulointerstitial nephritis* is the most common renal disease caused by Sjögren’s syndrome and may lead to *renal tubular acidosis (RTA)*, which in turn may cause *osteomalacia*. Even though the entire sequence chain was not directly observed in the data, the network could weave the individual relationships and create a more holistic view of the knowledge. Figure 3 presents a part of the network.

VII. Discussion

The long-term goal of this work is to create a nearly exhaustive list of factors about Sjögren’s syndrome by mining information from the medical literature. Like many other diseases, the factors associated with Sjögren’s syndrome can be categorized into four classes – “signs and symptoms”, “risk factors”, “associated condition” and “diagnostic tests”. In this paper, we specifically target two categories assuming that these two factors share causal relations with the disease and information about them in the text is represented using causal semantics. Our results confirm this assumption and show that by using causal relationships, we can extract “associated conditions” and “signs and symptoms” with better extraction performance than many other baseline sequence labeling models. Although some of the baseline systems, such as BioBERT and Gram-CNN, have been shown to perform well, a few factors explain why the performance was poor in this case. Firstly, these models are usually trained on larger datasets and typically contain 5,000 or more sentences [64]. These models were tested on standard tasks, such as disease, gene, protein identification, or specific relationships between them. Large benchmark datasets are already available for these tasks. Our case has no standard datasets, and we had to collect and annotate manually. Moreover, this task requires advanced knowledge in the domain area; annotation could only be done by users with prior experience working on Sjögren’s syndrome. As a result, annotation, in our case, was expensive and time-consuming. Building a large enough dataset for supervised models to perform well becomes impractical. Thus, an unsupervised or semi-supervised method is better suited to this context. Our method of detecting relationships, specifically causal relations, is a more scalable approach to address this problem.

Although we assume that causal relationships can be a helpful tool to retrieve two of the four labels, the present version of the causality extraction tool has some limitations, it assumes that there is only one relationship pair in the sentence. In reality, the sentences, particularly in scientific articles, are much more complex. One single sentence may have multiple relationships in multiple formats – triangular, i.e., two causes leading to one effect or one cause leading to two effects, transitive relations, and the presence of conditions

TABLE IV
SELECTED EXAMPLES OF EXTRACTING FACTORS BY MINING CAUSAL RELATIONSHIPS

	Sentence	Cause	Effect
1	Hypokalemic paralysis is a rare presentation of Fanconi syndrome (FS) caused by Sjogren's Syndrome.	Sjogren's Syndrome	Hypokalemic paralysis
2	Primary Sjogren's syndrome (pSS) is a chronic systemic autoimmune disease that leads to sicca symptoms, mainly xerophthalmia and xerostomia.	Primary Sjogren's syndrome	sicca symptoms, mainly xerophthalmia and xerostomia
3	sjogrens syndrome (SjS) is an autoimmune condition that primarily affects salivary and lacrimal glands, causing loss of secretion.	Sjogren's syndrome	loss of secretion
4	71-year-old woman in whom the diagnosis of possible causes of the development of annular erythema, led the team to identify primary Sjogren's syndrome (SS).	development of annular erythema	primary Sjogren's syndrome
5	Primary Sjogren's syndrome (pSS) is characterized by lymphocytic infiltration of the exocrine glands resulting in decreased saliva and tear production.	Primary Sjogrens Syndrome	decreased saliva and tear production
6	Development of non-Hodgkin's lymphoma (NHL) is the major adverse outcome of Sjogren's syndrome affecting both morbidity and mortality.	Sjogren's syndrome	non-Hodgkin's lymphoma

TABLE V
COMPARATIVE PERFORMANCE

Model	Precision	Recall	F1-score
Bi LSTM	0.45	0.84	0.59
Glove Embeddings + CNN	0.47	0.72	0.56
Bi LSTM + CRF	0.05	0.4	0.1
BioWordVec + CNN [65, 66]	0.48	0.74	0.58
BioBERT [64]	0.39	0.55	0.46
Gram-CNN [67]	0.52	0.74	0.61
Our approach	**0.85**	**0.78**	**0.81**

that deems the relationship true—for example, the sentence. "**sjogrens syndrome (SS)** is a rare condition characterized by **structural damage and secretory dysfunction of the lacrimal and salivary glands** that leads to **dryness, particularly xerophthalmia (eyes) and xerostomia (mouth)**." [3] demonstrates a transitive relation, and "**Sjogren's syndrome (SS)** is an autoimmune disease, among the most common ones, that targets mainly the **exocrine glands** as well as **extra-glandular epithelial tissues**." [4] has a triangular relation, where one event (Sjögren's syndrome) is causing two conditions. Identifying all such relations from a single sentence is beyond the scope of this work. As part of future work, we will address these limitations and build a more generic causal relationship extraction model that can extract multiple relationships from a single sentence, if present, furthermore, target inter-sentence causal relationships.

The results (Table V) show the central hypothesis of this work that causal relations can be used to extract certain factors associated with Sjögren's syndrome holds. It can retrieve several more factors from the article text compared to other baseline methods. However, on many occasions, *associated factors* or *signs and symptoms* are present in a sentence without any causal semantics. For example, the sentence "Two years after the presentation, the patient developed **dyspnea cough and xerostomia**" contains *symptoms*. However, due to the absence of a causal semantic, our present model will add this to the list of false negatives. To achieve the long-term goals and improve the recall of the model, it is essential to identify other relations that bind these factors with the disease. Similarly, for more accurate retrieval, we assume that "Sjögren's syndrome " will be present in the sentence and be part of the cause-effect pair; hence, the above sentence will not trigger our model for retrieval. This rationale for using other relations in the future will also help to extract the other two labels – "diagnostic tests" and "risk factors." As these two labels do not associate with the disease as a causality, we need to investigate the relations that will help to discover those factors. We will keep these tasks as part of the future directions of this work.

[3]https://pubmed.ncbi.nlm.nih.gov/28862467/
[4]https://pubmed.ncbi.nlm.nih.gov/29881381/

VIII. CONCLUSION AND FUTURE WORK

This paper presents an innovative approach to extracting factors related to Sjögren's syndrome from medical journal articles. We present a novel reinforcement learning-based method to identify causal relations from text and show that it outperforms most similar models. We apply this model to a dataset of 383 sentences from a more extensive set of 2,530 abstracts taken from articles on Sjögren's syndrome. Using causal relationships, we aimed to extract two out of four labels, "signs and symptoms" and "associated conditions," and show that our retrieval method has better precision, recall, and F1 scores than several supervised baseline models.

Although causal relations could effectively identify many such factors, several other types of relations bind the factors with Sjögren's syndrome. To improve the retrieval performance and cover the other two factors ("diagnostic test" and "risk factors"), as future directions, we will investigate other relations and build models that can identify and extract these labels from the text. Furthermore, we will improve our causal relationship extraction model to improve the coverage of relationship extraction and be able to extract multiple causal pairs from a single sentence, as well as discover inter-sentence relations.

Acknowledgment

This material is based upon work supported in part by the National Science Foundation under Grant No. 1948322.

References

[1] F. B. Vivino, "Sjogren's syndrome: Clinical aspects," Clinical Immunology, vol. 182, pp. 48–54, 2017, special issue: Sjogren's Syndrome.

[2] C. Q. Nguyen and A. B. Peck, "Unraveling the pathophysiology of sjogren syndrome-associated dry eye disease," The Ocular Surface, vol. 7, no. 1, pp. 11–27, 2009.

[3] C. G. Helmick, D. T. Felson, R. C. Lawrence, S. Gabriel, R. Hirsch, C. K. Kwoh, M. H. Liang, H. M. Kremers, M. D. Mayes, P. A. Merkel et al., "Estimates of the prevalence of arthritis and other rheumatic conditions in the united states: Part i," Arthritis & Rheumatism, vol. 58, no. 1, pp. 15–25, 2008.

[4] L. E. Brown, M. L. Frits, C. K. Iannaccone, M. E. Weinblatt, N. A. Shadick, and K. P. Liao, "Clinical characteristics of ra patients with secondary ss and association with joint damage," Rheumatology, vol. 54, no. 5, pp. 816–820, 2015.

[5] ——, "Clinical characteristics of RA patients with secondary SS and association with joint damage," Rheumatology, vol. 54, no. 5, pp. 816–820, 10 2014. [Online]. Available: https://doi.org/10.1093/rheumatology/keu400

[6] R. Patel and A. Shahane, "The epidemiology of sjögren's syndrome," Clinical epidemiology, vol. 6, p. 247, 2014.

[7] S. Ishikawa, T. Shoji, Y. Nishiyama, and K. Shinoda, "A case with acquired lacrimal fistula due to sjögren's syndrome," American Journal of Ophthalmology Case Reports, vol. 15, p. 100526, 2019.

[8] G. Nocturne and X. Mariette, "Advances in understanding the pathogenesis of primary sjögren's syndrome," Nature Reviews Rheumatology, vol. 9, no. 9, pp. 544–556, 2013.

[9] L. Dong, Y. Chen, Y. Masaki, T. Okazaki, and H. Umehara, "Possible mechanisms of lymphoma development in sjogren's syndrome," Current immunology reviews, vol. 9, no. 1, pp. 13–22, 2013.

[10] C. Lenfant, "Clinical research to clinical practice—lost in translation?" New England Journal of Medicine, vol. 349, no. 9, pp. 868–874, 2003.

[11] S. Bowman, G. Ibrahim, G. Holmes, J. Hamburger, and J. Ainsworth, "Estimating the prevalence among caucasian women of primary sjögren's syndrome in two general practices in birmingham, uk," Scandinavian journal of rheumatology, vol. 33, no. 1, pp. 39–43, 2004.

[12] E. Blanco, N. Castell, and D. I. Moldovan, "Causal relation extraction." in Lrec, 2008.

[13] A. Ittoo and G. Bouma, "Extracting explicit and implicit causal relations from sparse, domain-specific texts," in International Conference on Application of Natural Language to Information Systems. Springer, 2011, pp. 52–63.

[14] T. Dasgupta, R. Saha, L. Dey, and A. Naskar, "Automatic extraction of causal relations from text using linguistically informed deep neural networks," in 19th Annual SIGdial Meeting on Discourse and Dialogue, 2018, pp. 306–316.

[15] Q.-C. Bui, B. Ó. Nualláin, C. A. Boucher, and P. M. Sloot, "Extracting causal relations on hiv drug resistance from literature," BMC bioinformatics, vol. 11, no. 1, p. 101, 2010.

[16] M. Kyriakakis, I. Androutsopoulos, A. Saudabayev et al., "Transfer learning for causal sentence detection," 2019.

[17] Z. Li, Q. Li, X. Zou, and J. Ren, "Causality extraction based on self-attentive bilstm-crf with transferred embeddings," Neurocomputing, vol. 423, pp. 207–219, 2019.

[18] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," arXiv preprint arXiv:1312.5602, 2013.

[19] I. Hendrickx, S. N. Kim, Z. Kozareva, P. Nakov, D. Ó Séaghdha, S. Padó, M. Pennacchiotti, L. Romano, and S. Szpakowicz, "SemEval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals," in Proceedings of the 5th International Workshop on Semantic Evaluation. Uppsala, Sweden: Association for Computational Linguistics, Jul. 2010, pp. 33–38. [Online]. Available: https://www.aclweb.org/anthology/S10-1006

[20] H. Gurulingappa, A. M. Rajput, A. Roberts, J. Fluck, M. Hofmann-Apitius, and L. Toldo, "Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports," Journal of biomedical informatics, vol. 45, no. 5, pp. 885–892, 2012.

[21] M. Kyriakakis, I. Androutsopoulos, A. Saudabayev, and J. Ginés i Ametllé, "Transfer learning for causal sentence detection," in Proceedings of the 18th BioNLP Workshop and Shared Task. Florence, Italy: Association for Computational Linguistics, Aug. 2019, pp. 292–297. [Online]. Available: https://aclanthology.org/W19-5031

[22] H. Q. Yu, "Dynamic causality knowledge graph generation for supporting the chatbot healthcare system," in Proceedings of the Future Technologies Conference. Springer, 2020, pp. 30–45.

[23] M. S. Yin, M. Pomarlan, P. Haddawy, M. R. Tabassam, C. Chaimanakarn, N. Srimaneekarn, and S.-U. Hassan, "Automated extraction of causal relations from text for teaching surgical concepts," in 2020 IEEE International Conference on Healthcare Informatics (ICHI). IEEE, 2020, pp. 1–3.

[24] D. Chen, Y. Cao, and P. Luo, "Pairwise causality structure: towards nested causality mining on financial statements," in CCF International Conference on Natural Language Processing and Chinese Computing. Springer, 2020, pp. 725–737.

[25] C. Khoo, S. Chan, and Y. Niu, "The many facets of the cause-effect relation," in The Semantics of Relationships. Springer, 2002, pp. 51–70.

[26] J. Yang, S. C. Han, and J. Poon, "A survey on extraction of causal relations from natural language text," Knowledge and Information Systems, pp. 1–26, 2022.

[27] R. Girju, "Automatic detection of causal relations for question answering," in Proceedings of the ACL 2003 workshop on Multilingual summarization and question answering-Volume 12. Association for Computational Linguistics, 2003, pp. 76–83.

[28] C. S. G. Khoo, S. Chan, and Y. Niu, "Extracting causal knowledge from a medical database using graphical patterns," ser. ACL '00, 2000, pp. 336–343.

[29] C. Mihăilă and S. Ananiadou, "Semi-supervised learning of causal relations in biomedical scientific discourse," Biomedical engineering online, vol. 13, no. 2, pp. 1–24, 2014.

[30] D. Garcia et al., "Coatis, an nlp system to locate expressions of actions connected by causality links," in International Conference on Knowledge Engineering and Knowledge Management. Springer, 1997, pp. 347–352.

[31] R. Girju and D. I. Moldovan, "Text mining for causal relations," in Proceedings of the Fifteenth International Florida Artificial Intelligence Research Society Conference. AAAI Press, 2002, pp. 360–364. [Online]. Available: http://dl.acm.org/citation.cfm?id=646815.708596

[32] R. Mueller and S. Hüttemann, "Extracting causal claims from information systems papers with natural language processing for theory ontology learning," ser. Hawaii International Conference on System Sciences, 2018.

[33] S. Zhao, M. Jiang, M. Liu, B. Qin, and T. Liu, "Causaltriad: Toward pseudo causal relation discovery and hypotheses generation from medical text data," 2018.

[34] S. Sasaki, S. Takase, N. Inoue, N. Okazaki, and K. Inui, "Handling multiword expressions in causality estimation," in IWCS 2017—12th International Conference on Computational Semantics—Short papers, 2017.

[35] D. Kang, V. Gangal, A. Lu, Z. Chen, and E. Hovy, "Detecting and explaining causes from text for a time series event," arXiv preprint arXiv:1707.08852, 2017.
[36] D. Bollegala, S. Maskell, R. Sloane, J. Hajne, and M. Pirmohamed, "Causality patterns for detecting adverse drug reactions from social media: Text mining approach," JMIR public health and surveillance, vol. 4, no. 2, 2018.
[37] J. Chen, Q. Zhang, P. Liu, X. Qiu, and X. Huang, "Implicit discourse relation detection via a deep architecture with gated relevance network," in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Berlin, Germany: Association for Computational Linguistics, Aug. 2016, pp. 1726–1735. [Online]. Available: https://aclanthology.org/P16-1163
[38] E. M. Ponti and A. Korhonen, "Event-related features in feedforward neural networks contribute to identifying causal relations in discourse," in Proceedings of the 2nd Workshop on Linking Models of Lexical, Sentential and Discourse-level Semantics. Valencia, Spain: Association for Computational Linguistics, Apr. 2017, pp. 25–30. [Online]. Available: https://aclanthology.org/W17-0903
[39] Y. Xu, L. Mou, G. Li, Y. Chen, H. Peng, and Z. Jin, "Classifying relations via long short term memory networks along shortest dependency paths," in 2015 conference on empirical methods in natural language processing (EMNLP), 2015, pp. 1785–1794.
[40] Z. Li, Q. Li, X. Zou, and J. Ren, "Causality extraction based on self-attentive bilstm-crf with transferred embeddings," Neurocomputing, vol. 423, pp. 207–219, 2021.
[41] L. Wang, Z. Cao, G. De Melo, and Z. Liu, "Relation classification via multi-level attention cnns," in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2016, pp. 1298–1307.
[42] Y. Zhang, P. Qi, and C. D. Manning, "Graph convolution over pruned dependency trees improves relation extraction," arXiv preprint arXiv:1809.10185, 2018.
[43] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2018.
[44] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," in Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 2227–2237. [Online]. Available: https://aclanthology.org/N18-1202
[45] Y. Zhang, V. Zhong, D. Chen, G. Angeli, and C. D. Manning, "Position-aware attention and supervised data improve slot filling," in Conference on Empirical Methods in Natural Language Processing, 2017.
[46] S. Bowman, G. Ibrahim, G. Holmes, J. Hamburger, and J. Ainsworth, "Estimating the prevalence among caucasian women of primary sjögren's syndrome in two general practices in birmingham, uk," Scandinavian Journal of Rheumatology, vol. 33, no. 1, pp. 39–43, 2004. [Online]. Available: https://doi.org/10.1080/03009740310004676
[47] C. G. Helmick, D. T. Felson, R. C. Lawrence, S. Gabriel, R. Hirsch, C. K. Kwoh, M. H. Liang, H. M. Kremers, M. D. Mayes, P. A. Merkel, S. R. Pillemer, J. D. Reveille, J. H. Stone, and N. A. D. Workgroup, "Estimates of the prevalence of arthritis and other rheumatic conditions in the united states: Part i," Arthritis & Rheumatism, vol. 58, no. 1, pp. 15–25, 2008. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/art.23177
[48] M. M. Muñoz, J. Sebastián, R. Roda, Y. J. Soriano, and M. G. S. Pérez, "Sjögren's syndrome of the oral cavity. review and update," 2009.
[49] L. Douglas, "Facilitating timely diagnosis of sjögren's syndrome," BDJ Team, vol. 5, no. 2, p. 18026, 2018.
[50] S. G. Pasoto, V. A. de Oliveira Martins, and E. Bonfa, "Sjögren's syndrome and systemic lupus erythematosus: links and risks," Open access rheumatology: research and reviews, vol. 11, 2019.
[51] A. Rasmussen, L. Radfar, D. Lewis, K. Grundahl, D. U. Stone, C. E. Kaufman, N. L. Rhodus, B. Segal, D. J. Wallace, M. H. Weisman et al., "Previous diagnosis of sjögren's syndrome as rheumatoid arthritis or systemic lupus erythematosus," Rheumatology, vol. 55, no. 7, pp. 1195–1201, 2016.
[52] D. Marcheggiani and I. Titov, "Encoding sentences with graph convolutional networks for semantic role labeling," arXiv preprint arXiv:1703.04826, 2017.
[53] V. Konda and J. Tsitsiklis, "Actor-critic algorithms," Advances in neural information processing systems, vol. 12, 1999.
[54] D. Bahdanau, P. Brakel, K. Xu, A. Goyal, R. Lowe, J. Pineau, A. Courville, and Y. Bengio, "An actor-critic algorithm for sequence prediction," arXiv preprint arXiv:1607.07086, 2016.
[55] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "Albert: A lite bert for self-supervised learning of language representations," arXiv preprint arXiv:1909.11942, 2019.
[56] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?" arXiv preprint arXiv:1805.11604, 2018.
[57] H. Gurulingappa, A. Mateen-Rajpu, and L. Toldo, "Extraction of potential adverse drug events from medical case reports," Journal of biomedical semantics, vol. 3, no. 1, p. 15, 2012.
[58] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[59] Z. Huang, W. Xu, and K. Yu, "Bidirectional lstm-crf models for sequence tagging," arXiv preprint arXiv:1508.01991, 2015.
[60] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532–1543.
[61] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," ser. 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2018), 2018.
[62] J. Wang and W. Lu, "Two are better than one: Joint entity and relation extraction with table-sequence encoders," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online: Association for Computational Linguistics, Nov. 2020, pp. 1706–1721.
[63] S. Zhao, M. Hu, Z. Cai, and F. Liu, "Modeling dense cross-modal interactions for joint entity-relation extraction," in Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20, C. Bessiere, Ed. International Joint Conferences on Artificial Intelligence Organization, 7 2020, pp. 4032–4038, main track.
[64] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, and J. Kang, "Biobert: a pre-trained biomedical language representation model for biomedical text mining," Bioinformatics, vol. 36, no. 4, pp. 1234–1240, 2020.
[65] Y. Zhang, Q. Chen, Z. Yang, H. Lin, and Z. Lu, "Biowordvec, improving biomedical word embeddings with subword information and mesh," Scientific data, vol. 6, no. 1, pp. 1–9, 2019.
[66] A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou, and T. Mikolov, "Fasttext. zip: Compressing text classification models," arXiv preprint arXiv:1612.03651, 2016.
[67] Q. Zhu, X. Li, A. Conesa, and C. Pereira, "Gram-cnn: a deep learning approach with local context for named entity recognition in biomedical text," Bioinformatics, vol. 34, no. 9, pp. 1547–1554, 2018.