"Is Depression Related to Cannabis?": A Knowledge-Infused Model for Entity and Relation Extraction With Limited Supervision.

Kaushik Roy, Usha Lokala, Vedant Khandelwal and Amit Sheth

Artificial Intelligence Institute, University of South Carolina

Abstract

With strong marketing advocacy of the benefits of cannabis use for improved mental health, cannabis legalization is a priority among legislators. However, preliminary scientific research does not conclusively associate cannabis with improved mental health. In this study, we explore the relationship between depression and consumption of cannabis in a targeted social media corpus involving personal use of cannabis with the intent to derive its potential mental health benefit. We use tweets that contain an association among three categories annotated by domain experts - Reason, Effect, and Addiction. The state-of-the-art Natural Langauge Processing techniques fall short in extracting these relationships between cannabis phrases and the depression indicators. We seek to address the limitation by using domain knowledge; specifically, the Drug Abuse Ontology for addiction augmented with Diagnostic and Statistical Manual of Mental Disorders lexicons for mental health. Because of the lack of annotations due to the limited availability of the domain experts' time, we use supervised contrastive learning in conjunction with GPT-3 trained on a vast corpus to achieve improved performance even with limited supervision. Experimental results show that our method can significantly extract cannabis-depression relationships better than the state-of-the-art relation extractor. High-quality annotations can be provided using a nearest neighbor approach using the learned representations that can be used by the scientific community to understand the association between cannabis and depression better.

Keywords
Mental Health, Depression, Cannabis Crisis, Legalization, knowledge infusion, Relation Extraction

1. Introduction

Many states in the US have legalized the medical use of cannabis for therapeutic relief in those affected by Mental Illness [1] [2, 3]. The use of cannabis for depression, however, is not authorized yet [4]. Depression is ubiquitous among the US population, and some even use cannabis to self treat their depression[5][6]. Therefore, scientific research that can help understand the association between depression and cannabis consumption is of the utmost need given the fast increasing cases of depression in the US and consequent cannabis consumption [7].

Twitter can provide crucial contextual knowledge in understanding the usage patterns of cannabis consumption concerning depression [8, 9]. Conversations on social media such as Twitter provide unique insights as tweets are often unfiltered and honest in disclosing consumption patterns due to the anonymity and private space afforded to the users. For now, even with several platforms available to aid the analysis of depression concerning cannabis consumption, this understanding remains ambiguous [10, 11, 12]. Still, encouragingly there is support in the literature to show that cannabis use can be potentially associated with depressive patterns. Hence, we aim to extract this association as one of three relationships annotated by domain experts: Reason, Effect, and Addiction (Table 1).

Relationship	Tweet
Reason	"-Not saying im cured, but i feel less depressed lately, could be my #CBD oil supplement."
Effect	"-People will smoke weed and be on antidepressants. It's a clash!Weed is what is making you depressed."
Addiction	"-The lack of weed in my life is depression as hell."

Table 1
Cannabis-Depression Tweets and their relationships. Here the text in the blue and red represents the cannabis and depression entities respectively.

The paper studies mental health and its relationship with cannabis usage, which is a significant research problem. The study will help address several health challenges such "as the investigation of cannabis for the treatment of depression"," as a reason for depression" or "as an addictive phenomenon that accompanies depression". Extracting relationships between any *concepts/slang-terms/synonyms/street-names* related to '*cannabis,*' and '*depression,*' from text is a tough problem. This task is challenging for traditional Natural Language Processing (NLP) because of the immense variability with which tweets mentioning depression and cannabis are described. Here, we make use of the Drug Abuse Ontology (DAO) [13, 14] which is a domain-specific hierarchical framework containing 315 entities (814 instances) and 31 relations defining concepts about drug-abuse. The ontology has been used in prior work to analyze the effects of cannabis [15, 16, 17]. DAO was augmented using Patient Health Questionnaire 9th edition (PHQ-9), Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT), International Classificatio of Diseases 10th edition (ICD-10), Medical Subject Headings (MeSH) Terms, and Diagnostic and Statistical Manual for Mental Disorders (DSM-5) categories to infuse knowledge of mental health-related phrases in association with drugs such as cannabis [18][19]. Some of the triples extracted from the DAO are as follows: **(1)** SCRA → subClassOf → Cannabis; **(2)** Cannabis_Resin → has_slang_term → Kiff; **(3)** Marijuana_Flower → type → Natural_Cannabis.

For entity and relationship extraction (RE), previous approaches generally adopt deep learning models [20] [21]. However, these models require a high volume of annotated data and are hence unsuitable for our setting. Several pre-trained language representation models have recently advanced the state-of-the-art in various NLP tasks across various benchmarks [22, 23]. GPT-3 [24], BERT [25] are such language models [26]. Language models benefit from the abundant knowledge that they are trained on and, with minimal fine-tuning, can tremendously help

in downstream tasks under limited supervision. Hence, we exploit the representation from GPT-3 and employ supervised contrastive learning to deal with limited supervision in terms of quality annotations for the data. We propose a knowledge-infused deep learning framework based on GPT-3 and domain-specific DAO ontology to extract entities and their relationship. Then we further enhance the utilization of limited supervision through the use of supervised contrastive learning. It is well known that deep understanding requires many examples to generalize. Metric Learning frameworks such as Siamese networks have previously shown how limited supervision can help use contrastive learning with triplet loss [27]. Combinatorially this method leads to an increase in the number of training examples from an order of n to $nC3$, which helps with generalizability. The technique can also exploit the learned metric space representations to provide high-quality annotations over unlabeled data. Therefore, the combination of knowledge-infusion [28, 29, 30], pre-trained GPT-3, and supervised contrastive learning presents a very effective way to handle limited supervision. The proposed model has two modules: *(1) Phrase Extraction and Matching Module*, which utilizes the DAO ontology augmented with the PHQ-9, SNOMED-CT, ICD-10, MeSH Terms, and Diagnostic and Statistical Manual for Mental Disorders (DSM-5) lexicons to map the input word sequence to the entities mention in the ontology by computing the cosine similarity between the entity names (obtained from the DAO) and every n-gram token of the input sentence. This step identifies the depression and cannabis phrase in the sentence. Distributed representation obtained from GPT-3 of the depression phrase and cannabis phrase in the sentence is used to learn the contextualized syntactic and semantic information that complement each other. *(2) Supervised Contrastive Learning Module*, uses a triplet loss to learn a representation space for the cannabis and depression phrases through supervised contrastive learning. Phases with the correct relationship are trained to be closer in the learned representation space, and phrases with incorrect relationships are far apart.

Contributions:

(1) In collaboration with domain experts who provide limited supervision on real-world data extracted from Twitter, we learn a representation space to label the relationships between cannabis and depression entities to generate a cannabis-depression relationship dataset.

(2) We propose a knowledge-infused neural model to extract cannabis/depression entities and predict the relationship between those entities. We exploited domain-specific DAO ontology, which provides better coverage in entity extraction.

(3) Further, we use GPT-3 representations in a supervised contrastive learning approach to learn a representation space for the different cannabis and depression phrase relationships due to limited supervision.

(4) We evaluated our proposed model on the real world twitter dataset. The experimental results show that our model significantly outperforms the state-of-the-art relation extraction techniques by > 11% points on the F1 score.

1.1. Novel Contributions of the paper

(1) Semantic filtering: We use DAO, DSM-5, to extract contextual phrases expressed implicitly in the user tweet, mentioning Depression and Cannabis. This is required for noise-free domain adaptation of the model, as evident in our results.

(2) We develop a weak supervision pipeline to label the remaining 7000 tweets with three relationships (Reason, Effect, Addiction).

(3) We learn a domain-specific distance metric between the phrases, leveraging pre-trained GPT-3 embeddings of the extracted phrases and their relationship, in a supervised contrastive loss training setup.

(4) 7000 tweets were annotated and evaluated using the learned metric against the expert annotation with clustering (TSNE).

2. Related Works

Based on the techniques and their application to health, we discuss recent existing works. Standard DL approaches based on Convolutional Neural Networks (CNN), and Long Term Short Term Memory (LSTM) networks have been proposed for RE [31] [32, 33]. Hybrid models that combine CNN and LSTM have also been proposed [34]. More recently, Graph Convolutional Neural Networks (GCN)' s have been utilized to leverage additional structural information from dependency trees towards the RE task [35]. [36] guide the structure of the GCN by modifying the attention mechanism. Adversarial training has also been explored to extract entities and their relationships jointly [37]. Due to the variability in the specification of entities and relationships in natural language, [38, 39] have exploited entity position information in their DL framework. Models have demonstrated state of the art in RE based on the popular BERT language model, BioBERT [40], SciBERT [41], and XLNet Yang et al. [42]. Task-specific adaptations of BERT have been used to enhance RE in Shi and Lin [43] and Xue et al. [44]. Wang et al. [45] augment the BERT model with a structured prediction layer to predict multiple relations in one pass. In all the approaches discussed so far, knowledge has not been a component of the architecture [46].

Chan and Roth [47] show the importance of using knowledge to improve RE on sentences by showing an improvement of 3.9% of F1-score incorporating knowledge in an Integer Linear Programming (ILP) formulation.Wen et al. [48] use the attention weights between entities to guide traversal paths in a knowledge graph to assist RE. Distiawan et al. [49] use knowledge graph TransE embeddings in their approach to improving performance. Some of the other prominent work utilizing knowledge graph for relation extraction is [50, 51, 52].

These methods, however, do not consider a setting in which the availability of high-quality annotated data is scarce. We use knowledge to extract relevant parts of the sentence [53, 54] and pre-trained GPT-3 representations trained over a massive corpus in conjunction with supervised contrastive learning to achieve a high degree of sample efficiency with limited supervision.

3. Our Approach

3.1. Dataset

The dataset we have used for our study consists of 11,000 Tweets collected using the twitris API from Jan 2017 to Feb 2019 - determined by three substance use epidemiologists as a pe-

riod of heightened Cannabis consumption. The experts annotated 3000 tweets (due to time constraints) with one of 3 relationships that they considered essential to identify: "Reason," "Effect," and "Addiction." The annotation had a Cohen Kappa Agreement of 0.8. Example from each of these different relationships already shown in Table 1

3.2. Phrase Extraction and Matching

We exploit the domain-specific knowledge base to replace the phrases in social media text with the knowledge base concepts under this method. The Phrase Extraction and Matching are performed in several steps, which are:

- **Depression and cannabis Lexicon:** We have exploited the state of the art Drug Abuse Ontology (DAO) to extract various medial entities and slang terms related to cannabis and depression. We further expand the entities using entities extracted from PHQ-9, SNOMED-CT, ICD-10, MeSH Terms, and Diagnostic and Statistical Manual for Mental Disorders (DSM-5).

- **Extracting N-Grams from Tweets:** The N-Grams are extracted from the tweets are considered to better understand the context of the terms by taking into consideration the words around it. For example, from the tweet *whole world emotionally depressed everybody needs smoke blunt to relax We living nigga*, we will obtain ngrams such as whole world, emotionally depressed, depressed everybody, need smoke, need smoke blunt, living nigga.

- **GPT-3:**Generative Pre-Trained Transformer 3 is an autoregressive language model. We have used GPT-3 to generate the embedding of the N-Grams generated and the cannabis and depression Phrase because of the vast dataset it is trained on, which provides us the phrases' embeddings based on its global understanding. Cosine Similarity: It is a measure of similarity between two non-zero vectors in a similar vector space. This metric is often used to get a semantic similarity between two phrase embeddings obtained in the same vector space.

- **Phrase Matching:** We use the cosine similarity metric to understand the semantic similarity between the phrases. We have taken a threshold of 0.75 or more cosine similarity. Once the phrase obtains a similarity value more than or equals the threshold, the original N-Grams from the tweet text are replaced by the matched cannabis/depression Phrase. The above steps are repeated for all the tweets. For example, we would obtain *emotionally depressed* as the depression phrase, whereas *need smoke blunt* is found to be the cannabis phrase.

3.3. Supervised Contrastive Learning

The proposed model architecture is divided into two sub-parts: A) Transformer Block, B) Loss Function. The input tweet is first sent through a block of 12 transformers, and later the embedding is passed through a triplet loss function. The label associated with the tweet input is
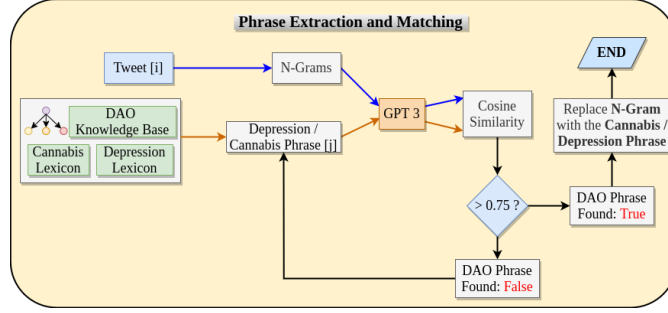
Figure 1: The Phrase extraction and Mapping Pipeline, where the DAO is used to extract GPT-3 representations of the related phrases

used to extract a complimentary sample (a tweet with the same label) and a negative sample (a tweet associated with a different label). These positive and negative samples are sent through a block of 12 transformers to obtain their embedding, which is further passed on to the triplet loss function. Under the loss, the function tries to achieve a low cosine similarity between the tweet and its negative sample as close to 0. At the same time, it tries to achieve a high cosine similarity between the tweet and its positive sample as close to 1.

$$CoSim(A, P) - CoSim(A, N) + \alpha \leq 0 \qquad (1)$$

Where A is the anchor (the initial data point), P is a positive data point which is of the same class as the anchor, N is a negative data point which is of the different class as the anchor. CoSim(X, Y) is the cosine similarity between the two data points, and α is the margin. For the example shown in Section 3.2, if we consider the anchor sample to be **"whole world emotionally depressed everybody needs smoke blunt relax We living nigga"**, corresponding to that positive sample is *"Depressionarmy weed amp sleep I awake I depressed"* and the negative sample would be *"This weird rant like weed makes anxiety depression worse Im soooo sick ppl like jus"*.
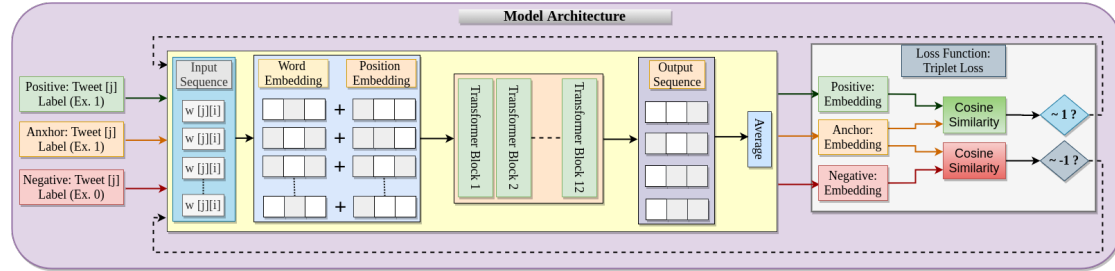

Figure 2: Supervised Constrastive Learning pipeline using Triplet Loss

4. Experimental Setup and Results

In this section, we discuss the results of the task of cannabis and depression RE tasks. After that, we will also provide a technical interpretation of the results.

4.1. Results

The dataset utilized in our experiment is described in Section-3 [55] . We have used Precision, Recall, and F1 score as the metric to compare our proposed methodology with the state-of-the-art relation extractor. As a baseline model, we used BERT, BioBERT, and its various variations such as [56]:

- **$BERT_{PE}$**: We expand the BERT as a position embedding along with the BERT embedding with the position data (the relative distance of the word with cannabis and depression entity) obtained via domain-specific knowledge resource.

- **$BERT_{PE+PA}$**: This consists of an additional component of position-aware mechanism along with $BERT_{PE}$.

Table-2 Summarises the performance of our model over the baselines. Our proposed methodology has outperformed all the state of the art models based on the given metrics. As compared to the worst-performing baseline, BioBERT our model achieves an F1 score with an absolute improvement of 12.32%, and with best performing baseline $BERT_{PE+PA}$ it gives an improvement of 11.28% in F1 Score. From the above comparison using contrastive learning with knowledge, infused learning can perform better in relation classification.

Method	Precision	Recall	F1-Score
BERT	64.49	63.22	63.85
BioBERT	63.97	62.15	63.06
$BERT_{PE}$	60.64	56.51	58.50
$BERT_{PE+PA}$	65.41	65.25	64.50
Proposed Model	74.6	76.17	75.37

Table 2
Performance comparison of our proposed model with baselines methods

4.2. Ablation Study

We have performed the ablation study by removing one component from the proposed model, evaluate its performance to understand the impact of various components. Based on the study, we found that, as we remove the contrastive loss function from our learning approach, the model performance significantly drops by 6.46 % F1 Score, 6.53 % Recall, and 6.4 % Precision. The significant decrease in the model's performance shows that generating an embedding for two samples of the same class similar and of different classes dissimilar brings in a great advantage to the training of the model. The contrastive loss function allows us to learn the representation of the same classes to be closer to each other in vector space and hence allows us in generating the representation of unlabelled data from the dataset (discussed further, later in this section)

We also observe that domain-specific knowledge resource with contextualized embedding trained over a large corpus (GPT-3) is very important. As we further remove the second component from our model, we see a total decrease of 9.01 % F1 score, 8.92 % Precision, and 9.11 %

recall in the proposed model's performance. This component (Knowledge Infusion) was majorly responsible for removing the data's ambiguity using the phrases from human-curated domain-specific knowledge bases (such as DAO, DSM-5, SNOMED-CT, and others). Also, the contextualized embedding helped us consider the global representation of the entities present in the dataset and hence contribute to improving the model's performance.

Model	Precision	Recall	F1-Score
Proposed Model	74.6	76.17	75.37
(-) Contrastive learning loss	68.2 ($\downarrow$8.5%)	69.64 ($\downarrow$8.6%)	68.91 ($\downarrow$8.57%)
(-) knowledge infusion	65.68 ($\downarrow$11.9%)	67.06 ($\downarrow$11.96%)	66.35 ($\downarrow$11.97%)

Table 3
Ablation Study over the proposed model to evaluate the effect of contrastive learning loss and knowledge infusion in determining the relationship between cannabis and depression

Thus, this shows that every component of the proposed model is necessary for the best performing results.

4.3. Cluster Analysis

After training the model, we annotate the unlabelled data in the dataset by classifying among the three relationships. We parse the unlabelled tweets from the first module to extract the phrases from the knowledge bases using contextualized embedding. Later the embeddings are pushed into the proposed model architecture to obtain a representation of the tweets. The representation is used to create a cluster of the tweet data points and determine the label of the un-labeled tweets based on the majority of the data points present. The representation of the cluster after labeling unlabelled tweets is shown in Figure 3. Some examples from each of the cluster are as follows:

- **Reason:** 1) *Depressionarmy weed amp sleep I awake I depressed,* 2) *mood smoke blunt except the fact I depressed,* 3) *weed hits ya RIGHT depression,* 4) *I smoked weed drank alcohol drowning sorrows away,* 5) *whole world emotionally depressed everybody need smoke blunt relax We livin nigga*

- **Effect:** 1) *marijuana bad marijuana makes feel depressed low mmk,* 2) *Unemployed stoners are the most depressed on the planet,* 3) *guess depression took long time discover marijuana makes VERY DEPRESSED alcohol doesnt help either,* 4) *This weird rant like weed makes anxiety depression worse Im soooo sick ppl like jus,* 5) *waking weed naps nigga feeling depressed hell*

- **Addiction:** 1) *I feel like weed calm someone suffer depression anxiety psychosis predisposed either,* 2) *Small trigger warning Blaine suffers anxiety depression occasionally smoke pot,* 3) *need blunt accompany depression,* 4) *This bot crippling depression ok weed lol,* 5) *Violate blunt distraction possibly despair bask commitment This would never happen*
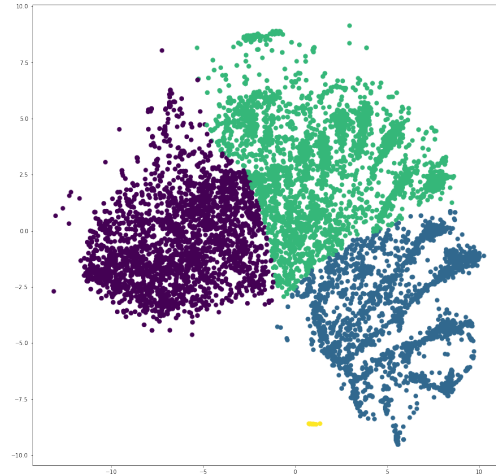
Figure 3: Shows the depression and cannabis phrases grouped according to the relationships of Reason, Effect and Addiction. Purple - Reason, Green - Effect, Blue and a little yellow below - Addiction.

5. Reproducibility

From this study, we will be delivering high quality annotated dataset of 3000 tweets along with the full annotated dataset (by our method) of 11000 tweets, will be made publicly available [1] to support research on the psychological impact of cannabis and depression use during COVID-19, as Cannabis use related to depression is seeing a rise once more. Also, the trained model will be shared for reproducibility of the results and annotation of tweets. Unfortunately, we cannot release the code used for training at this time as recently, Microsoft bought the rights to the GPT-3 model. Therefore, to use the learning method proposed in this paper, GPT-3 will need to be substituted with an alternative language model such as BERT, GPT-2, etc. [2]

6. Conclusion

In this study, we present a method to determine the relationship between depression and cannabis consumption. We motivate the necessity of understanding this issue due to the rapid increase in cases of depression in the US and across the world and subsequent increase in cannabis consumption. We utilize tweets to understand the relationship as tweets are typically unfiltered expressions of simple usage patterns among cannabis users who use it in association with their depressive moods or disorder. We present a knowledge aware method to determine the relationship significantly better than the state-of-the-art effectively, show the quality of the learned relationship through visualization on TSNE based clusters, and annotate the unlabeled parts of the dataset. We show by training on this new dataset (human-labeled and estimated label) that the model's prediction quality is improved. We present this high-quality dataset for utilization by the broader scientific community in better understanding the relationship

---

[1]https://github.com/khvedant02/A-knowledge-infused-model-for-Entity-and-Relation-Extraction-with-Limited-Supervision
[2]https://blogs.microsoft.com/blog/2020/09/22/microsoft-teams-up-with-openai-to-exclusively-license-gpt-3-language-model/

between depression and cannabis consumption.

7. Broader Impact

Although we develop our method to handle relationship extraction between depression and cannabis consumption specifically, we generally develop a domain knowledge infused relationship extraction mechanism that uses state-of-the-art language models, few shot machine learning techniques (contrastive loss) to achieve efficient and knowledge guided extraction. We see the improved quality in the results over transformer models. We believe that for applications with real-life consequences such as these, it is crucial to infuse domain knowledge as a human would combine with language understanding obtained from language models to identify relationships efficiently. Humans typically can learn from very few examples. Motivated by this and the lack of availability of examples, we develop our relation extraction method. We hope our significantly improved results will encourage scientists to explore further the use of domain knowledge infusion in application settings that demand highly specialized domain expertise.

References

[1] K. Hanson, A. Garcia, State medical marijuana laws, in: National Conference of State Legislatures, 2014.

[2] R. Room, Legalizing a market for cannabis for pleasure: Colorado, washington, uruguay and beyond, 2014.

[3] N. D. Volkow, R. D. Baler, W. M. Compton, S. R. B. Weiss, Adverse health effects of marijuana use, N. Engl. J. Med. 370 (2014) 2219–2227.

[4] M. B. Bridgeman, D. T. Abazia, Medicinal cannabis: History, pharmacology, and implications for the acute care setting, P T 42 (2017) 180–188.

[5] A. S. Young, R. Klap, R. Shoai, K. B. Wells, Persistent depression and anxiety in the united states: prevalence and quality of care, Psychiatr. Serv. 59 (2008) 1391–1398.

[6] S. E. Lankenau, J. Ataiants, S. Mohanty, S. Schrager, E. Iverson, C. F. Wong, Health conditions and motivations for marijuana use among young adult medical marijuana patients and non-patient marijuana users, Drug Alcohol Rev. 37 (2018) 237–246.

[7] K. M. Keyes, M. Wall, M. Cerdá, J. Schulenberg, P. M. O'Malley, S. Galea, T. Feng, D. S. Hasin, How does state marijuana policy affect US youth? medical marijuana laws, marijuana use and perceived harmfulness: 1991-2014, Addiction 111 (2016) 2187–2195.

[8] O. Corazza, S. Assi, P. Simonato, J. Corkery, F. S. Bersani, Z. Demetrovics, J. Stair, S. Fergus, C. Pezzolesi, M. Pasinetti, P. Deluca, C. Drummond, Z. Davey, U. Blaszko, J. Moskalewicz, B. Mervo, L. D. Furia, M. Farre, L. Flesland, A. Pisarska, H. Shapiro, H. Siemann, A. Skutle, E. Sferrazza, M. Torrens, F. Sambola, P. van der Kreeft, N. Scherbaum, F. Schifano, Promoting innovation and excellence to face the rapid diffusion of novel psychoactive substances in the EU: the outcomes of the ReDNet project, Hum. Psychopharmacol. 28 (2013) 317–323.

[9] L. Burns, A. Roxburgh, R. Bruno, J. Van Buskirk, Monitoring drug markets in the internet

age and the evolution of drug monitoring systems in australia, Drug Test. Anal. 6 (2014) 840–845.

[10] P. A. Cavazos-Rehg, K. Zewdie, M. J. Krauss, S. J. Sowles, "no high like a brownie high": A content analysis of edible marijuana tweets, Am. J. Health Promot. 32 (2018) 880–886.

[11] R. Daniulaityte, F. R. Lamy, G. A. Smith, R. W. Nahhas, R. G. Carlson, K. Thirunarayan, S. S. Martins, E. W. Boyer, A. Sheth, "retweet to pass the blunt": Analyzing geographic and content features of Cannabis-Related tweeting across the united states, J. Stud. Alcohol Drugs 78 (2017) 910–915.

[12] F. R. Lamy, R. Daniulaityte, M. Zatreh, R. W. Nahhas, A. Sheth, S. S. Martins, E. W. Boyer, R. G. Carlson, "you got to love rosin: Solventless dabs, pure, clean, natural medicine." exploring twitter data on emerging trends in rosin tech marijuana concentrates, Drug Alcohol Depend. 183 (2018) 248–252.

[13] D. Cameron, G. A. Smith, R. Daniulaityte, A. P. Sheth, D. Dave, L. Chen, G. Anand, R. Carlson, K. Z. Watkins, R. Falck, PREDOSE: a semantic web platform for drug abuse epidemiology using social media, J. Biomed. Inform. 46 (2013) 985–997.

[14] U. Lokala, R. Daniulaityte, F. Lamy, M. Gaur, K. Thirunarayan, U. Kursuncu, A. P. Sheth, Dao: An ontology for substance use epidemiology on social media and dark web, JMIR Public Health and Surveillance (2020).

[15] U. Lokala, F. R. Lamy, R. Daniulaityte, A. Sheth, R. W. Nahhas, J. I. Roden, S. Yadav, R. G. Carlson, Global trends, local harms: availability of fentanyl-type drugs on the dark web and accidental overdoses in ohio, Comput. Math. Organ. Theory (2018) 1–12.

[16] D. Cameron, G. A. Smith, R. Daniulaityte, A. P. Sheth, D. Dave, L. Chen, G. Anand, R. Carlson, K. Z. Watkins, R. Falck, Predose: a semantic web platform for drug abuse epidemiology using social media, Journal of biomedical informatics 46 (2013) 985–997.

[17] U. Kursuncu, M. Gaur, U. Lokala, A. Illendula, K. Thirunarayan, R. Daniulaityte, A. Sheth, I. B. Arpinar, What's ur type? contextualized classification of user types in marijuana-related communications using compositional multiview embedding, in: 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), IEEE, 2018, pp. 474–479.

[18] M. Gaur, U. Kursuncu, A. Alambo, A. Sheth, R. Daniulaityte, K. Thirunarayan, J. Pathak, Let me tell you about your mental health!: Contextualized classification of reddit posts to DSM-5 for web-based intervention, in: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 2018, pp. 753–762.

[19] M. Gaur, A. Alambo, J. P. Sain, U. Kursuncu, K. Thirunarayan, R. Kavuluru, A. Sheth, R. Welton, J. Pathak, Knowledge-aware assessment of severity of suicide risk for early intervention, in: The World Wide Web Conference, 2019, pp. 514–525.

[20] Y. Lin, S. Shen, Z. Liu, H. Luan, M. Sun, Neural relation extraction with selective attention over instances, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2016, pp. 2124–2133.

[21] J. Lee, S. Seo, Y. S. Choi, Semantic relation classification via bidirectional lstm networks with entity-aware attention using latent entity typing, Symmetry 11 (2019) 785.

[22] J. Maillard, S. Clark, D. Yogatama, Jointly learning sentence embeddings and syntax with unsupervised tree-lstms, Natural Language Engineering 25 (2019) 433–449.

[23] A. Akbik, T. Bergmann, R. Vollgraf, Pooled contextualized embeddings for named entity

recognition, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019, pp. 724–728.

[24] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., Language models are few-shot learners, arXiv preprint arXiv:2005.14165 (2020).

[25] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186. URL: https://www.aclweb.org/anthology/N19-1423. doi:10.18653/v1/N19-1423.

[26] C. Lin, T. Miller, D. Dligach, S. Bethard, G. Savova, A bert-based universal model for both within-and cross-sentence clinical temporal relation extraction, in: Proceedings of the 2nd Clinical Natural Language Processing Workshop, 2019, pp. 65–71.

[27] E. Hoffer, N. Ailon, Deep metric learning using triplet network, in: International Workshop on Similarity-Based Pattern Recognition, Springer, 2015, pp. 84–92.

[28] A. Sheth, M. Gaur, U. Kursuncu, R. Wickramarachchi, Shades of knowledge-infused learning for enhancing deep learning, IEEE Internet Computing 23 (2019) 54–63.

[29] U. Kursuncu, M. Gaur, A. Sheth, Knowledge infused learning (k-il): Towards deep incorporation of knowledge in deep learning, arXiv preprint arXiv:1912.00512 (2019).

[30] M. Gaur, U. Kursuncu, A. Sheth, R. Wickramarachchi, S. Yadav, Knowledge-infused deep learning, in: Proceedings of the 31st ACM Conference on Hypertext and Social Media, 2020, pp. 309–310.

[31] C. Liu, W. Sun, W. Chao, W. Che, Convolution neural network for relation extraction, in: International Conference on Advanced Data Mining and Applications, Springer, 2013, pp. 231–242.

[32] M. Miwa, M. Bansal, End-to-end relation extraction using lstms on sequences and tree structures, arXiv preprint arXiv:1601.00770 (2016).

[33] S. Yadav, A. Ekbal, S. Saha, A. Kumar, P. Bhattacharyya, Feature assisted stacked attentive shortest dependency path based bi-lstm model for protein–protein interaction, Knowledge-Based Systems 166 (2019) 18–29.

[34] D. Liang, W. Xu, Y. Zhao, Combining word-level and character-level representations for relation classification of informal text, in: Proceedings of the 2nd Workshop on Representation Learning for NLP, 2017, pp. 43–47.

[35] F. Wu, A. Souza, T. Zhang, C. Fifty, T. Yu, K. Weinberger, Simplifying graph convolutional networks, in: K. Chaudhuri, R. Salakhutdinov (Eds.), Proceedings of the 36th International Conference on Machine Learning, volume 97 of *Proceedings of Machine Learning Research*, PMLR, Long Beach, California, USA, 2019, pp. 6861–6871. URL: http://proceedings.mlr.press/v97/wu19e.html.

[36] Z. Guo, Y. Zhang, W. Lu, Attention guided graph convolutional networks for relation extraction, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Florence, Italy, 2019, pp. 241–251. URL: https://www.aclweb.org/anthology/P19-1024. doi:10.18653/

`v1/P19-1024`.

[37] G. Bekoulis, J. Deleu, T. Demeester, C. Develder, Adversarial training for multi-context joint entity and relation extraction, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Brussels, Belgium, 2018, pp. 2830–2836. URL: https://www.aclweb.org/anthology/D18-1307. doi:`10.18653/v1/D18-1307`.

[38] S.-P. Choi, Extraction of protein–protein interactions (ppis) from the literature by deep convolutional neural networks with various feature embeddings, Journal of Information Science 44 (2018) 60–73.

[39] Y. Peng, Z. Lu, Deep learning for extracting protein-protein interactions from biomedical literature, arXiv preprint arXiv:1706.01556 (2017).

[40] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, J. Kang, Biobert: pre-trained biomedical language representation model for biomedical text mining, arXiv preprint arXiv:1901.08746 (2019).

[41] I. Beltagy, K. Lo, A. Cohan, Scibert: A pretrained language model for scientific text, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 3606–3611.

[42] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, Q. V. Le, Xlnet: Generalized autoregressive pretraining for language understanding, arXiv preprint arXiv:1906.08237 (2019).

[43] P. Shi, J. Lin, Simple bert models for relation extraction and semantic role labeling, arXiv preprint arXiv:1904.05255 (2019).

[44] K. Xue, Y. Zhou, Z. Ma, T. Ruan, H. Zhang, P. He, Fine-tuning bert for joint entity and relation extraction in chinese medical text, arXiv preprint arXiv:1908.07721 (2019).

[45] H. Wang, M. Tan, M. Yu, S. Chang, D. Wang, K. Xu, X. Guo, S. Potdar, Extracting multiple-relations in one-pass with pre-trained transformers, arXiv preprint arXiv:1902.01030 (2019).

[46] M. Gaur, K. Faldu, A. Sheth, Semantics of the black-box: Can knowledge graphs help make deep learning systems more interpretable and explainable?, arXiv preprint arXiv:2010.08660 (2020).

[47] Y. S. Chan, D. Roth, Exploiting background knowledge for relation extraction, Proceedings of the 23rd International Conference on (2010).

[48] D. Wen, Y. Liu, K. Yuan, S. Si, Y. Shen, Attention-Aware Path-Based relation extraction for medical knowledge graph, in: Smart Computing and Communication, Springer International Publishing, 2018, pp. 321–331.

[49] B. Distiawan, G. Weikum, J. Qi, R. Zhang, Neural relation extraction for knowledge base enrichment, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 229–240.

[50] J. Li, G. Huang, J. Chen, Y. Wang, Dual cnn for relation extraction with knowledge-based attention and word embeddings, Computational intelligence and neuroscience 2019 (2019).

[51] H. Zhou, C. Lang, Z. Liu, S. Ning, Y. Lin, L. Du, Knowledge-guided convolutional networks for chemical-disease relation extraction, BMC bioinformatics 20 (2019) 260.

[52] P. Li, K. Mao, X. Yang, Q. Li, Improving relation extraction with knowledge-attention, arXiv preprint arXiv:1910.02724 (2019).

[53] A. Alambo, M. Gaur, U. Lokala, U. Kursuncu, K. Thirunarayan, A. Gyrard, A. Sheth, R. S. Welton, J. Pathak, Question answering for suicide risk assessment using reddit, in: 2019 IEEE 13th International Conference on Semantic Computing (ICSC), IEEE, 2019, pp. 468–473.

[54] U. Kursuncu, M. Gaur, C. Castillo, A. Alambo, K. Thirunarayan, V. Shalin, D. Achilov, I. B. Arpinar, A. Sheth, Modeling islamist extremist communications on social media using contextual dimensions: religion, ideology, and hate, Proceedings of the ACM on Human-Computer Interaction 3 (2019) 1–22.

[55] U. Lokala, R. Daniulaityte, R. Carlson, F. Lamy, S. Yadav, A. Sheth, Social media data for exploring the association between cannabis use and depression, https://doi.org/10.6084/m9.figshare.14067122.v2, 2021. Accessed: 2019-2-28.

[56] S. Yadav, U. Lokala, R. Daniulaityte, K. Thirunarayan, F. Lamy, A. Sheth, " when they say weed causes depression, but it's your fav antidepressant": Knowledge-aware attention framework for relationship extraction, arXiv preprint arXiv:2009.10155 (2020).