Received 10 May 2023, accepted 28 May 2023, date of publication 31 May 2023, date of current version 13 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3281544

RESEARCH ARTICLE

Emotion Classification in Texts Over Graph Neural Networks: Semantic Representation is Better Than Syntactic

IQRA AMEER[1], NECVA BÖLÜCÜ[2,3], GRIGORI SIDOROV[4], AND BURCU CAN[5]

[1]Computer Science Program, Division of Science and Engineering, Pennsylvania State University at Abington, Abington, PA 19001, USA
[2]Department of Computer Engineering, Hacettepe University, 06800 Ankara, Turkey
[3]Data61, CSIRO, Sydney, NSW 2060, Australia
[4]Instituto Politécnico Nacional, Centro de Investigación en Computación, Mexico City 07738, Mexico
[5]Department of Computing Science and Mathematics, University of Stirling, FK9 4LA Stirling, U.K.

Corresponding author: Grigori Sidorov (sidorov@cic.ipn.mx)

This work was supported in part by the Mexican Government, Mexico, through Consejo Nacional de Ciencia y Tecnología (CONACYT), under Grant A1-S-47854; and in part by Secretaría de Investigación y Posgrado of the Instituto Politécnico Nacional, Mexico, under Grant 20220852 and Grant 20220859.

ABSTRACT Social media is a widely used platform that provides a huge amount of user-generated content that can be processed to extract information about users' emotions. This has numerous benefits, such as understanding how individuals feel about certain news or events. It can be challenging to categorize emotions from text created on social media, especially when trying to identify several different emotions from a short text length, as in a multi-label classification problem. Most previous work on emotion classification has focused on deep neural networks such as Convolutional Neural Networks and Recurrent Neural Networks. However, none of these networks have used semantic and syntactic knowledge to classify multiple emotions from a text. In this study, semantic and syntactic aware graph attention networks were proposed to classify emotions from a text with multiple labels. We integrated semantic information in the graph attention network in the form of Universal Conceptual Cognitive Annotation and syntactic information in the form of dependency trees. Our extensive experimental results showed that our two models, UCCA-GAT (accuracy = 71.2) and Dep-GAT (accuracy = 68.7), were able to outperform the state-of-the-art performance on both the challenging SemEval-2018 E-c: Detecting Emotions (multi-label classification) English dataset (accuracy = 58.8) and GoEmotions dataset (accuracy = 65.9).

INDEX TERMS Emotion classification, GAT, UCCA, dependency, semantic, syntactic, social media.

I. INTRODUCTION

Emotions are defined by Hwang and Matsumoto [1] as innate constructs typically produced throughout socializing and aid in interpersonal interaction, which is a significant part of daily life. Emotions fundamentally influence human life, which affects our decisions and mental and physical health [2]. There are positive and negative emotions; positive emotions are more associated with improving human health as well as work efficiency, while negative emotions may cause health problems. Emotions can be observed from two broader perspectives: neurological and human felt experiences [3]. Ekman [4] identified 6 basic emotions: anger, disgust, joy, fear, sadness, and surprise. Ekman's basic emotions are the result of research on facial expressions. There are also other popular theories of emotions, such as Plutchik's wheel of emotions [5] and Parrots' classification of emotions [6]. Plutchik [5] used a wheel for emotions in which the center is basic (core) emotions of sadness, disgust, anger, anticipation, joy, fear, surprise, and trust, and radiating toward the outer edges are less intense variants of these core emotions. Parrot's classification of emotions [6] uses a tree-structured list in which the first level consists of six basic emotions.

Humans tend to express their emotions through different channels. In particular, social media is a popular platform where individuals express their emotions in different forms,

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar.

TABLE 1. Multi-label instances for emotion classification taken from different datasets.

ID	Text	Emotions
SemEval-2018Task-1C		
1	In dire need of a worship session to just let go of everything	anticipation, fear, pessimism, sadness
2	@wabermes The @RavalliRepublic had a good one but then the reporter quit	neutral
3	Whatever you decide to do make sure it makes you #happy.	joy, love, optimism
GoEmotions		
1	Thats great! Just curious, what game?	joy, surprise
2	To make her feel threatened	fear
3	Unfortunately, I don't. Sorry.	anger, sadness

such as text, image, audio, or video. This study is based on textual emotion classification, where the goal of emotion classification, an extended field of sentiment analysis, is to assign possible emotions to a piece of text that most accurately reflect the mental state of the author. There are three ways to solve the emotion classification problem based on the approach: (1) binary emotion classification detects whether an emotion is present or not [7], (2) multi-class emotion classification classifies an instance into one of the predefined set of n labels [8], [9], (3) multi-label emotion classification classifies a given instance as "neutral or no emotion" or one or more from a set of predefined n labels that best represent the mental state of the author [10], [11]. The development of emotion classification models is crucial, given their widespread influence and presence. There are numerous applications for emotion categorization models in various domains, including financial marketing [12], [13], [14], medicine [15], [16], [17], education [18], [19], [20], etc. Therefore, emotion classification in the text is a well-studied task in Natural Language Processing (NLP). There are various multi-label classification datasets with different numbers of emotion labels, such as GoEmotions [21] has 28 emotions (28th emotion label is neutral), SemEval-2018 Task-1C [10] a shared task dataset contains 12 emotions (12th emotion label is neutral). Since the number of emotion labels in the GoEmotions [21] dataset is higher than in other datasets, resulting in an imbalance dataset, the studies [22], [23], [24] use the GoEmotions dataset with emotions of Ekman [4]. In this study, we also mapped the emotion labels of the GoEmotions dataset to Ekman's basic 6 labels (see IV-A2). Table 1 represents instances of both the SemEval-2018 and GoEmotions datasets.

With the domination of deep learning models in NLP tasks such as sentiment analysis, question answering, and machine translation, wildly used deep learning models are also applied for multi-label classification [11], [25], including Convolutional Neural Networks (CNN) [26], Long Short-term Memory (LSTM) networks [27], Bi-directional Long Short-term Memory (Bi-LSTM) networks [28], Gated Recurrent Units (GRUs) [29], Attention [30], and Multi-head Attention [31], [32]. Graph neural networks [33] are a class of deep learning methods designed to perform inference on data described by graphs and have been extensively employed in various NLP tasks, such as dependency parsing [34], [35], text classification [36], [37]. The models commonly neglect the semantic meaning of the text, concerning meaning defined by relations between words in a sentence, which is essential for classifying the emotions of a given text.

Semantic representation reflects the meaning of the text in a rather structured form (e.g., graph-based or tree-based representation) [38]. In recent years, graph-based representations gained researchers' attention due to their ability to express and generate adequate target structures, especially for a text's sentence-level syntactic analysis and semantic representation. The increasing popularity of graph-based semantic representations has led to the proposal of various semantic representation frameworks such as Abstract Meaning Representation [39], Universal Conceptual Cognitive Annotation [40], bilexical Semantic Dependencies [41], Universal Decompositional Semantics [42], and Parallel Meaning Bank [43]. These graph-based representations have proven to be beneficial in Natural Language Understanding (NLU) tasks and have already demonstrated their applicability in a variety of NLP tasks such as text summarisation [44], [45], paraphrase detection [46], [47], machine translation [48], [49], question answering [50], [51], and text simplification [52].

In this study, we proposed semantically and syntactically aware models to investigate the impact of incorporating semantic and syntactic representation with Graph Attention Network (GAT) for multi-label classification. In order to incorporate the semantic information, we constructed the GAT model by integrating the UCCA-based semantic representation of the expressions. As a syntactic representation, we utilized dependency trees of the expressions. The overview of the proposed model is illustrated in Figure 1. First, we extracted UCCA-based semantic representations and dependency trees from the multi-label classification datasets. Then, we extracted adjacency and feature matrices from semantic and syntactic representations, respectively. Finally, we use the matrices in the proposed GAT model for multi-label classification. The detailed steps are described in section III.

The contribution of this study is as follows:

- We proposed a semantic and syntactic aware GAT for the multi-label classification problem. To the best of our knowledge, this is the first study to integrate the UCCA framework into GATs for the multi-label classification task and compare the effects of semantic and syntactic representation for the problem.

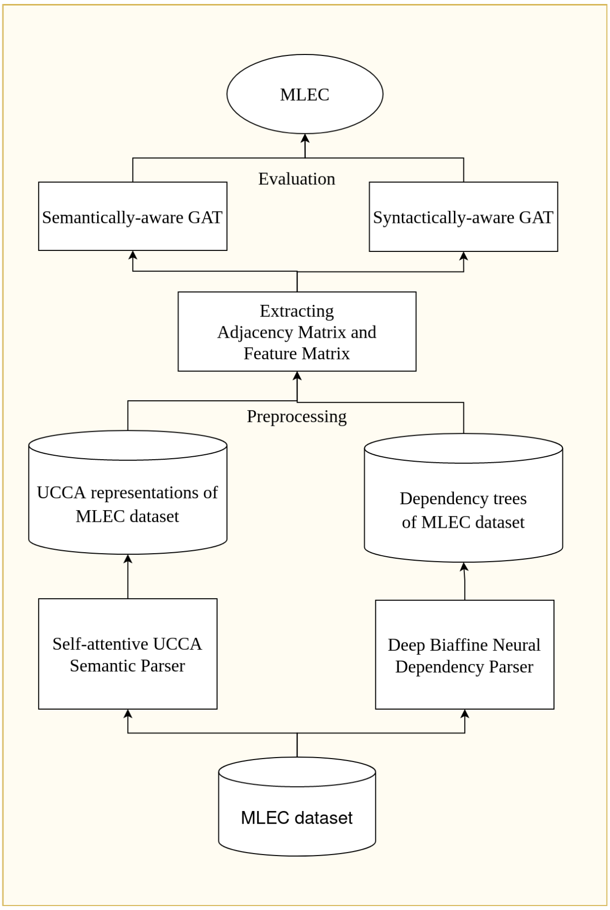


FIGURE 1. The overview of the proposed system.

- We evaluated our proposed model on the multi-label classification datasets, namely the challenging multi-label classification dataset of SemEval-2018 E-c: Detecting Emotions (multi-label classification) [10] and GoEmotions [21] with Ekman emotions [4].
- We compared semantically and syntactically aware models in detail for the multi-label classification problem. The experimental results showed that the semantically-aware model outperforms the syntactically-aware model on SemEval-2018 E-c: Detecting Emotions and GoEmotions datasets. Moreover, the semantically-aware model outperforms the state-of-the-art performance on both datasets.

The rest of the paper is organized as follows. Section II reviews related work on the multi-label classification problem and semantic and syntactic aware studies proposed for NLP problems. Section III describes our methodology for addressing the multi-label classification problem with details on semantic and syntactic parser models and preprocessing methodology. Section IV presents our experimental setup, results, and detailed analysis of the results. Finally, Section V concludes the paper with insights on the impact of semantic and syntactic information on the multi-label classification problem and possible future work.

II. LITERATURE REVIEW

Emotion detection has been studied in recent years [53], [54], [55], and with the success of deep learning models in NLP tasks, such as Neural Machine Translation (NMT) [48], [56], Semantic Textual Similarity (STS) [57], [58], the advanced deep learning models also employed in solving the multi-label emotion classification problem, such as LSTM [59], [60], CNN [61], GRU [62], [63], Transformers [11], [25], [64]. The multi-label emotion classification problem is a popular task that is also addressed in the SemEval-2018 shared task [10] for English, Arabic, and Spanish. The SemEval shared tasks released for emotion classification play an important role in developing resources for emotion classification [10], [59]. NTUA-SLP [59], which took first place in the emotion classification subtask of the SemEval-2018 shared task [10], presented BiLSTM with a multi-layer self-attention mechanism. As pre-trained embedding, they trained word2vec [65] with 550 million English tweets, augmented with a set of effective word embeddings trained with the word embeddings. To fine-tune the hyperparameters of the proposed model, the authors adopted a Bayesian optimization approach that allows for a time-efficient search for all possible values in high-dimensional space instead of a grid or random search. One of the other teams that participated in SemEval-2018 is TCS Research [66], which combined three different features (lexicon, deep learning features extracted from BiLSTM, SentiNeuron [67] features) into an SVM to develop a unified ensemble system. The system was designed to handle noisy sentiment multi-label datasets with a mixture of embeddings in parallel. The team that placed third in the SemEval shared task [10], PlusEmo2Vec, applied neural network models to extract features and used those features in traditional machine learning classifiers (logistic regression and support vector regression (SVR)). They also extended the training set by using an external dataset provided by the competition to learn a better representation of emojis and #hashtags. The other studies also applied DL networks, such as MLP [68], LSTM [60].

In addition to the SemEval shared task [10], multi-label emotion detection [21], [69], [70], [71] is a well-studied problem. Ameer et al. [71] applied pre-trained language models for MLEC problem, such as XLNet, DistilBERT, and RoBERTa with multiple attention mechanisms, and fine-tuned them on a SemEval-2018 E-c dataset for English and Ren-CECps dataset for Chinese. The proposed models outperformed the state-of-the-art results on both datasets. Li and Xiao [72] developed a multi-label emotion recognition tool called Multi-EmoBERT. They applied it to SemEval2018 Task 1 and several fake news and corpora, examining the relationships between veracity/stance and emotion and achieving state-of-the-art performance. Farruque et al. [73] analyzed multiple emotions in tweets by utilizing multi-label classifiers to identify basic emotions and those specific to depression. They used a hybrid emotion model consisting of common emotions from four psychological models of emotions and added new emotion categories that are significant for examining depression. The research findings indicate that the Deep Learning model outperformed the RankSVM algorithm in modeling the complex semantic features of the new emotional categories. Islam et al. [74] used two approaches to MLEC: problem transformation and algorithm adaptation. The results showed that binary relevance and label powerset methods perform better than other multi-label classifiers. The random forest classifier is better than the support vector machine as a base classifier for problem transformation methods. The paper also shows that SenticNet[1] can improve the accuracy of the models.

In recent years, semantically and syntactically aware models have gained popularity due to impressive performance in NLP problems [75], [76], [77], [78], [79], [80].

Semantically aware studies that concern the semantic relations between words of a group of words in a sentence can be divided into models that use semantic similarity [75] and semantic information, such as graph-based semantic frameworks [76], [78], which are utilized in different NLP problems, i.e., text classification problems (irony detection [75], content detection [76]), reading comprehension [78] and machine translation [79], [80]. One of the semantically-aware models is SemBERT (semantics-aware BERT) [81], which uses PropBank [82] for semantic role label sequences. The model takes the raw text sequences and the semantic role label sequences as embedding vectors to feed a pre-trained BERT [83]. The model is performed on ten NLU benchmark datasets involving natural language inference, machine reading comprehension, semantic similarity, and text classification. Ek et al. [84] investigated the effect of syntactic and semantic representations with LSTM for the language model. They trained the LSTM language model on sentences annotated with universal syntactic dependency roles [85], dependency tree features, and universal semantic tags [43]. Elbasani and Kim [76] proposed a neural approach for toxic content detection. The study is based on a CNN model that integrates an AMR graph-based semantic representation as the input layer of the model. Nguyen et al. [86] presented a graph embedding algorithm using AMR and compared various well-known machine learning systems, such as Seq2seq, Conv2Seq, and Transformer by integrating the AMR graph embedding representation. Slobodkin et al. [87] presented two novel encoders called Scene-Aware Self-Attention (SASA) and Scene-Aware Cross-Attention (SACRA) that integrate UCCA graph-based semantic representation into Transformer for machine translation.

Syntactically aware models have also effectively solved various NLP problems, i.e., sentiment analysis, text generation [88], [89], question answering [90], and semantic role labeling [90], etc. In particular, dependency trees are used as syntactic information in problems such as machine translation [91], [92], language model [93], [94], and semantic role labeling [95]. Bastings et al. [91] presented Graph Convolutional Network (GCN) using a dependency tree for machine translation. Nguyen et al. [96] proposed a hierarchical accumulation tree structure using dependency trees as a self-attention mechanism for machine translation. Using graph-attention network on a dependency tree structure and external pre-training knowledge from the BERT language model, Huang et al. [88] used a graph attention network on a dependency tree structure and external pre-training knowledge from the BERT language model [88] to better describe the relationship between context and aspect words. The dependency tree graphs included the BERT subwords, allowing a more accurate representation of words by graph attention. Three datasets are used in experiments to show how well the model works.

Guo et al. [89] treated text generation as a graph generation problem that takes advantage of word order and syntactic linkages. The method involved incremental sentence construction while maintaining syntactic integrity using a top-down approach based on syntax. Experimental results on both synthetic and real text generation tasks demonstrated the effectiveness of the proposed approach. Schlichtkrull et al. [90] worked on question answering and semantic role labeling problems by developing a graph neural network. The authors developed a post-hoc technique for analyzing GNN predictions highlighting irrelevant edges. They showed that such a classifier might be trained in a completely differentiable manner with stochastic gates and the expected L0 norm, which promotes sparsity and uses an attribution method to analyze GNN models for the above two tasks. It provided insight into the information flow in these models and revealed that a significant portion of edges could be omitted without negatively affecting the model's performance. The remaining edges can then be analyzed to interpret the model predictions. Marcheggiani and Titov [95] used GCN to encode the constituent structures and provide the information for the semantic role labeling system. The SpanGCN nodes in the proposed technique corresponded to the components. The initial node representations are created in the first stage by "composing" the first and last words of the constituent into word representations. In the second stage, graph convolutions are created based on the constituent tree, leading to syntactically informed constituent representations. Then, the semantic role labeling classifier fed the "decomposed" constituent representations " back into word representations. The SpanGCN was compared to alternatives, such as a model that favors dependency trees over GCNs, and its effectiveness was demonstrated using the English SRL benchmarks CoNLL-2005 [97], CoNLL-2012 [98], and FrameNet [99].

In summary, semantically and syntactically aware models are famous in NLP problems because they significantly outperform naive learning models [100]. Inspired by these models, we proposed a graph neural network for the multi-label classification that employed semantic and syntactic information. To the best of our knowledge, none of the prior studies

[1]https://sentic.net/ Last visited: 04-05-2023

have explored semantically and syntactically aware models for multi-label classification.

III. METHODOLOGY

In this section, we described the proposed architecture for multi-label classification with the preprocessing steps.

The methodology consists of two steps:

1) **Extracting representation**: As input to the graph attention network, we used semantic and syntactic representations to compare the problem; for this, we used external semantic and dependency parser models with similar structures for each sample in the datasets.
2) **Applying proposed model**: We proposed a graph attention network to categorize each sample for emotions using the representations extracted in the first step.

A. BACKGROUND

This study explored representations over graph neural networks for the multi-label classification problem. We used a graph-based semantic representation called Universal Conceptual Cognitive Annotation (UCCA) and a syntactic representation (dependency tree) to extract graphs for the multi-label classification.

1) SEMANTIC PARSING

Semantic representation is a way of expressing the meaning of a text that a machine can process to serve a particular NLP task that requires meaning understanding [101]. Recently, NLP problems have been using semantic representations such as text summarization [44], [45], question answering [102], or machine translation [48]. We utilized semantic representation for the multi-label classification problem and chose the UCCA [40] graph-based semantic representation for the study.

UCCA [40] is a proposed graph-based semantic representation with a multi-layered framework where each layer corresponds to a ''module'' of semantic distinction. The foundational layer of UCCA is represented by a directed acyclic graph (DAG), where the nodes represent terminal (words) and non-terminal tokens, and the edges represent the semantic roles between the nodes. In the UCCA representation, there are 4 different categories of semantic roles:

1) **Scene Elements**: The main element of the UCCA representation is the Scene and Process (P), State (S), Participant (A), and Adverbial (D) are the elements of a Scene. Process (P) and State (S) are the main relations of a Scene, which determine the type of the Scene. If there is an action or movement, the main relation is Process (P). However, if it is a temporally persistent state, the relation is State (S). Participant (A) is the participant of the main relation, and there may be one or more participants in the Scene. The last element Adverbial (D), describes the main relation in detail, e.g., time or location.
2) **Non-scene Elements**: There are also elements of the UCCA that do not evoke a Scene. The elements are Center (C), Elaborator (E), Connector (N), and Relator (R), where Center (C) and Elaborator E) conceptualize the non-Scene unit and also act as class descriptors. Connector (N) connects two or more entities with similar roles, e.g., Elaborator or Center. Finally, Relator (R) relates one or more entities to the main Scene.
3) **Inter-scene Relations**: UCCA may contain more than one Scene, called Parallel Scene (H). Linker L) is an inter-scene relation that connects Parallel Scenes. Ground (G) is an entity that refers to the speech event of the speaker, the hearer in which the text was uttered/written/conceived.
4) **Other**: Function (F) is an element that functions as an auxiliary to a larger construction, e.g. tense or focus.

We applied a semantic parser proposed by Bolucu and Can [103] as an external semantic parsing model to extract the semantic representation of the two datasets. The parser model is a graph-based semantic parser that solves the problem as a constituency parsing problem. It comprises an encoder and decoder where the encoder is a self-attention mechanism of Vaswani et al. [31] with 2 MLP classifiers with 2 fully-connected layers and a nonlinear activation function ReLU as the output layer. The output layer generates per-span scores where spans correspond to the constituents in the constituency tree. The decoder part is CYK (Cocke-Younger-Kasami) [104] algorithm that generates a constituency tree with a maximum score using the scores generated in the output layer of the encoder. The model transforms the constituency trees into UCCA representations by using one of the MLP layers to predict the remote edges of the representation.

2) DEPENDENCY PARSING

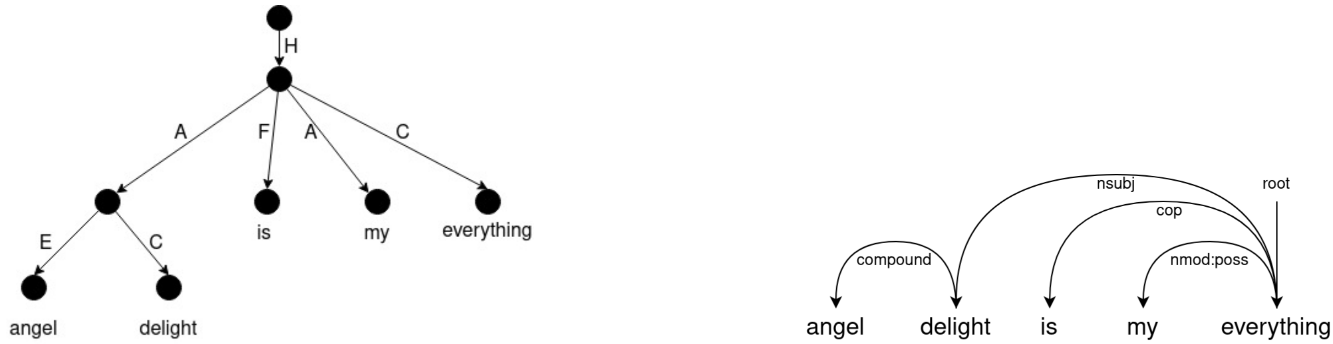
Dependency grammar is an approach to the syntax of natural languages. Dependency is the notion of linguistic units that are words connected to each other by directed links [105].

We applied a dependency parser proposed by Dozat and Manning [106] called Deep Biaffine Neural Dependency Parser to extract the syntactic representation of the datasets for the problem. The parser model follows the Bi-LSTM model with biaffine classifiers to predict arcs and labels. An example of a UCCA representation (Figure 2a) and a dependency tree (Figure 2b) from the SemEval-2018 dataset is presented in Figure 2.

B. PRE-PROCESSING

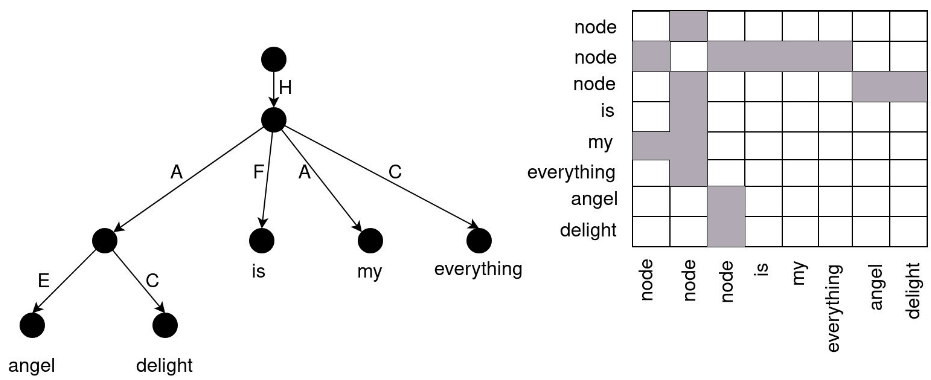
We proposed a graph attention network to solve the problem of classifying emotions with multiple labels. We used semantic and syntactic representations as input to the model to extract the adjacency matrix and the feature matrix from graphs.

UCCA is an acyclic directed graph $G = (V, E)$, where V is a set of nodes that are terminal and non-terminal nodes in the UCCA representation, and E is a set of edges

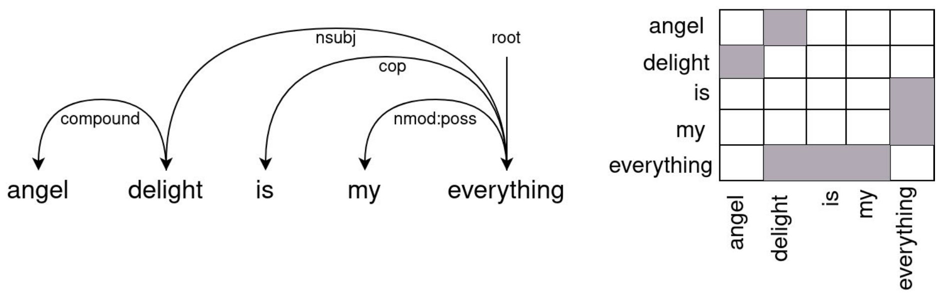


(a) Semantic Representation (b) Syntactic rRepresentation

FIGURE 2. **The semantic and syntactic representations of the Tweet "angel delight is my everything" taken from the SemEval-2018 dataset obtained from the parser models, i.e., the semantic parser and dependency parser.**



(a) Semantic Adjacency Matrix



(b) Syntactic Adjacency Matrix

FIGURE 3. **The adjacency matrix extracted from semantic and syntactic representations of the Tweet "angel delight is my everything" taken from the SemEval-2018 dataset. The gray color in the matrix represents the value of 1 and the white color to the value of 0.**

that are UCCA semantic roles. We extracted the feature matrix X ($n \times k$, where n is the number of nodes (terminal and non-terminal) in the graph, and k is the embedding dimension) and the adjacency matrix A ($n \times n$, where n is the number of nodes in the graph) from the UCCA graph. For the feature matrix, pre-trained word embeddings (BERT [83], RoBERTa [107], etc.) were used for terminal nodes, and a randomly generated embedding with the same dimension of the pre-trained embedding was used for non-terminal nodes. The same procedure was applied to the dependency trees extracted from the dependency parser.

Figure 3 demonstrates examples of adjacency matrices extracted from semantic (Figure 3a) and syntactic (Figure 3b) representations. All preprocessing is illustrated in Figure 4.

We only used the nodes of representations. Since the representations of the texts in the dataset are of different sizes, we used the pad function of the numpy [108] library to scale the adjacency and feature matrices and bring them to the same size.

C. PROPOSED METHOD

GAT is a novel neural network architecture that operates on graph-structured data [109]. The model leverages masked

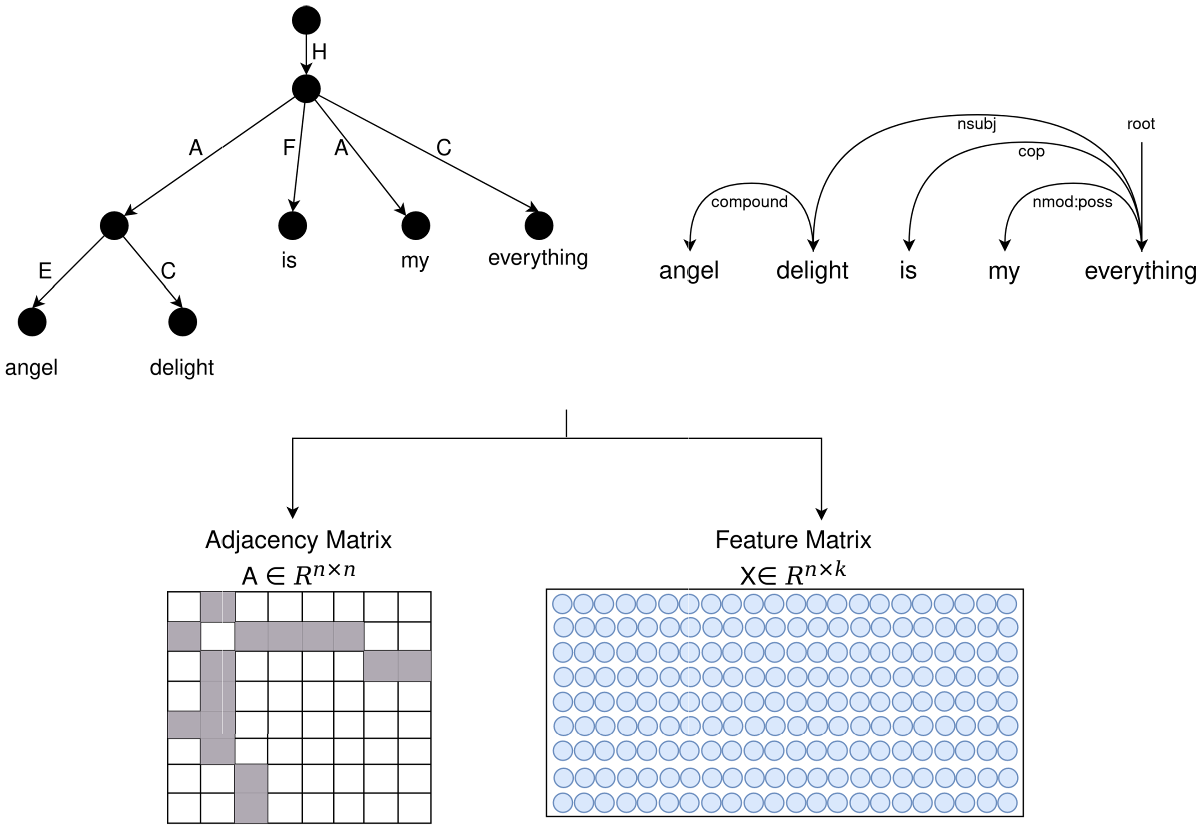


FIGURE 4. **Preprocessing for graph attention network.**

self-attentional layers of [31] to address the limitations of prior methods based on graph convolutions by adding attention to each neighbour [110].

The model is composed of three layers: (1) the input layer, (2) the self-attentional layer, and (3) the output layer. The architecture of the proposed model is demonstrated in Figure 5.

- **Input Layer**: The input layer of the model is designed as an adjacency and feature matrix that is generated from the either semantic representation of the dependency tree of the sample.
- **Graph Attention Layer**: We applied self-attentional layer of [31] that is computed as:

$$H^{i+1} = \sigma \left(A \cdot H^{i} \cdot W^{i} \right) \tag{1}$$

  where Wi is the weight matrix for layer i, A is the adjacency matrix, H^i is the feature matrix of the first layer ($H^0 = X$), where X is feature matrix (extracted in preprocessing step), and σ is the ReLU non-linear activation function. In the proposed model, we applied a multi-layer GAT where the layer size is a hyperparameter that needs to be tuned in the graph.
- **Output Layer**: The output layer is the sigmoid layer with m classes, where m is the number of emotions in the dataset. The sigmoid layer squeezes the results between 0 and 1, and we used 0.5 as the threshold to convert the probabilities into classes. The equation 2 presents the formula of the layer.

$$Z = sigmoid(H^{1}) \tag{2}$$

  where H^l is the feature matrix of the final graph attention layer.

IV. EXPERIMENTS AND RESULTS

This section presents the experimental setup, including datasets, external paring models (semantic and syntactic), evaluation measures, hyperparameter settings, and the results and analysis.

A. EXPERIMENTAL SETUP

1) TRAINING DETAILS OF EXTERNAL PARSERS

- **Semantic Parser**: The semantic parser is trained with the combination of all training sets of English, French, and German datasets released in the SemEval-2019 shared task [111]. The model uses the concatenation of word embeddings and syntactic embeddings (PoS tag, dependency label, entity type, and entity IOB) as input. We used Stanza library[2] [112] to extract syntactic features of the multi-label classification datasets.

[2]https://stanfordnlp.github.io/stanza/ Last visited: 17-08-2022

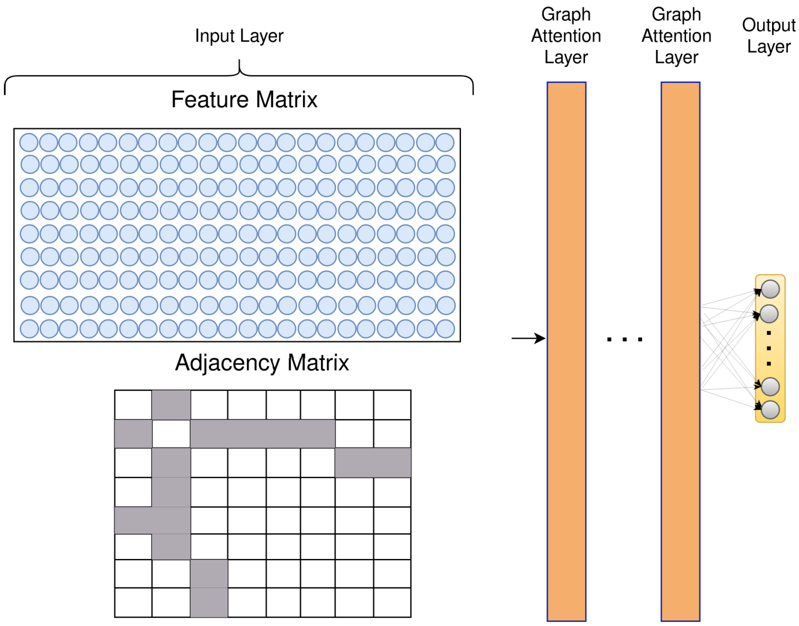


FIGURE 5. Architecture of the proposed model.

- **Syntactic Parser**: The model is trained with Universal Dependencies v2.5 treebanks [85].

2) DATASET

We performed extensive experimentation on two multi-label emotion classification datasets, i.e., SemEval-2018 Task-1C and GoEmotions. The details of the datasets are given below:

- SemEval-2018 Task-1C: Affect in Tweets dataset contains Tweets collected from 2016 to 2017, and it was developed for Task E-c: Detecting Emotions (multi-label classification) shared task [10]. This dataset is annotated with the presence/absence of 11 emotions (anger, anticipation, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, and trust). The dataset is provided in three languages (Arabic, Spanish, and English) for the emotion classification task. However, we only used English language data for this study.
- GoEmotions: A Dataset of Fine-Grained Emotions contains 58k samples taken from Reddit comments [21]. This dataset is annotated with the 28 emotions (27 emotion categories or neutral). We mapped 28 labels to Ekman's [4] 6 categories (anger (anger, annoyance, disapproval), disgust (disgust), fear (fear, nervousness), joy (admiration, amusement, approval, caring, desire, excitement, gratitude, joy, optimism, pride, relief), sadness (disappointment, embarrassment, grief, remorse, sadness) and neutral (neutral)). We randomly split the dataset into train (80%), development (10%), and test (10%) sets as defined in the paper [21].

The statistics of the datasets are provided in Table 2 (see Appendix I for detailed statistics).

TABLE 2. Dataset statistics.

Dataset	Train Set	Dev Set	Test Set	Total
SemEval-2018 Task-1C	6,838	886	3,259	10,983
GoEmotions	43,410	5,426	5,427	54,263

3) EVALUATION MEASURE

Since this is a multi-label classification challenge, each piece of text can have one or more gold emotion labels and one or more predicted emotion labels. In this study, we used multi-label accuracy (a.k.a. Jaccard index), the size of the intersection of the predicted labels, and the gold labels divided by the size of the union of the predicted and gold labels. This evaluation is performed for each piece of text t in the test dataset and then averaged over all instances in dataset D.

$$Accuracy = \frac{1}{|D|}\sum_{t\in D}\frac{G_t \cap P_t}{G_t \cup P_t} \tag{3}$$

where G_t is the set of gold labels for sentence s, P_s is the set of predicted labels for sentence t, and D is the total number of sentences in the test set.

In addition to the Jaccard index (multi-label accuracy), we also reported Macro-averaged F-score and the Micro-averaged F-score[3] [25].

4) HYPERPARAMETER SETTINGS

We implemented the proposed model using the PyTorch library [113]. The Adam optimizer [114] was used with an epsilon value of $1e-8$ and default max grad norm.

[3]https://competitions.codalab.org/competitions/17751. Last visited: 17-08-2022

TABLE 3. Emotion classification results obtained from SemEval-2018 Task-1C dataset.

Models	Accuracy	Micro F_1	Macro F_1
Our proposed models			
UCCA-GAT	**61.2**	66.1	60.0
Dep-GAT	59.7	63.5	57.8
Top three state-of-the-art models			
NTUA-SLP [59]	58.8	70.1	52.8
TCS Research [67]	58.2	69.3	53.0
PlusEmo2Vec [116]	57.6	69.2	49.7

TABLE 4. Emotion classification results obtained from GoEmotions dataset.

Models	Accuracy	Micro F_1	Macro F_1
Our proposed models			
UCCA-GAT	**71.2**	75.4	63.9
Dep-GAT	68.7	74.7	61.1
Top three state-of-the-art models			
BERT [21]	-	-	64.0
RoBERTa [71]	65.9	69.1	61.8
Dim-RoBERTa [71]	65.7	68.6	61.0

For nodes corresponding to words, we used pre-trained language models (BERT [83], RoBERTa [107], etc.) in the feature matrix. The model was fine-tuned using the development set of datasets (see Appendix II for detailed hyperparameters).

B. RESULTS AND ANALYSIS

1) RESULTS

Tables 3 and 6 represent the multi-label accuracy, Micro F_1, and Macro F_1 results obtained by applying the proposed semantically and syntactically aware model on the SemEval-2018 Task-1C [10] and GoEmotions [21] datasets, respectively. In these Tables, "Models" refers to the two proposed (UCCA-GAT and Dep-GAT) models and state-of-the-art studies on both datasets to compare the results. The "UCCA-GAT" model refers to a GAT model with an input layer consisting of adjacency and feature matrices extracted from the UCCA semantic representation. Similarly, the "Dep-GAT" model also uses a GAT but with an input layer consisting of adjacency and feature matrices extracted from dependency trees.

The results show that semantically and syntactically aware models are the most suitable for the multi-label classification problem on two different natures of the texts (Tweets and Reddit comments). On SemEval-2018 Task-1C [10] dataset, overall, the best results are obtained with the semantically aware UCCA-GAT model (accuracy = 61.2). Our proposed models, UCCA-GAT and Dep-GAT, outperformed the top three state-of-the-art approaches on the SemEval-2018 Task-1C dataset. On GoEmotions [21] dataset, overall, the semantically aware UCCA-GAT model performed best (accuracy = 71.2). Our proposed models, UCCA-GAT and Dep-GAT, outperformed the top three state-of-the-art studies on the GoEmotions dataset.

To understand the reason for the better results of the semantically aware model (UCCA-GAT), we analyzed the adjacency and feature matrices extracted from semantic and dependency parsers. As explained in Section III-A, the UCCA representation is a graph with nodes that are terminal (words) and non-terminal (multi-words). The non-terminal nodes, which represent the semantics of multi-tokens, increase the depth of the graphs compared to the dependency trees. This also increases the density of the adjacency matrix extracted from the UCCA semantic representation and helps to better performance in multi-label classification problem.

TABLE 5. Scores for each emotion in SemEval-2018 Task-1C dataset.

	UCCA-GAT			Dep-GAT		
Emotion	Precision	Recall	F1	Precision	Recall	F1
Anger	85.1	81.8	83.4	75.6	72.7	74.1
Anticipation	35.1	63.8	45.3	26.9	52.0	35.5
Disgust	73.6	89.0	80.6	74.8	88.0	80.8
Fear	32.3	76.3	45.4	28.6	72.1	41.0
Joy	77.6	91.2	83.9	82.5	65.3	72.9
Love	51.2	75.7	61.1	50.4	73.3	59.7
Optimism	67.8	85.2	75.5	70.4	73.5	71.9
Pessimism	81.3	60.4	69.3	58.8	76.0	66.3
Sadness	73.4	77.1	75.2	64.3	71.4	67.7
Surprise	52.9	15.2	23.6	23.1	52.9	32.1
Trust	54.2	09.4	16.0	21.7	54.2	31.0

TABLE 6. Scores for each emotion in GoEmotions dataset.

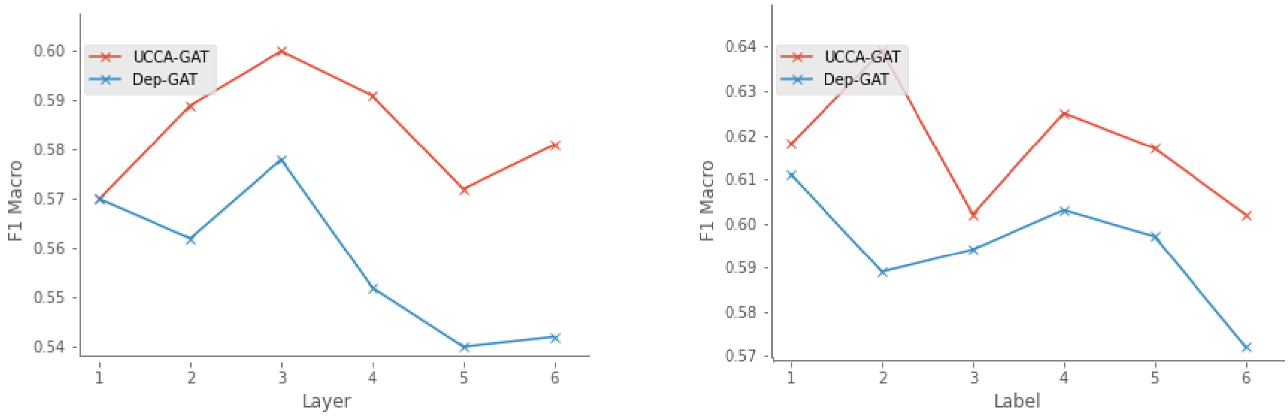
	UCCA-GAT			Dep-GAT		
Emotion	Precision	Recall	F1	Precision	Recall	F1
Anger	85.8	84.0	84.9	90.7	68.9	78.3
Disgust	32.9	81.3	46.8	27.6	78.9	40.9
Fear	15.2	73.5	25.2	11.3	56.1	18.9
Joy	88.9	76.2	82.0	84.9	86.0	85.4
Sadness	32.0	68.0	43.5	41.1	55.4	47.2
Surprise	75.0	88.6	81.2	64.2	95.3	76.8
Neutral	79.7	87.9	83.6	82.8	78.3	80.5

To understand the performance of the models for each emotion, we computed the precision, recall, and macro F_1 scores of the best performing model (UCCA-GAT) for the SemEval-2018 Task-1C and GoEmotions datasets. The scores are presented in Table 5 and 6 for the SemEval-2018 Task-1C and GoEmotions datasets, respectively.

The results show that both semantically and syntactically aware models performed better on emotions "anger", "disgust," "joy," "pessimism," and "sadness" on the SemEval-2018 Task-1C dataset and emotions "anger", "disgust," "joy," "optimism" on GoEmotions dataset. One possible reason could be the percentages of instances of these particular emotions in the datasets (see Table 8).

To summarize, we obtained comparatively higher results on the GoEmotions dataset. The following are possible reasons for the lower performance on the SemEval-2018 Task-1C dataset:

- The number of emotion classes in the datasets (SemEval-2018 Task-1C: 11 and GoEmotions: 7) could lead to the low performance of the proposed model on the SemEval-2018 Task-1C dataset.
- The experimental results of the parsers show that the models perform better on shorter text than longer ones



(a) SemEval-2018 Task-1C dataset

(b) GoEmotions dataset

FIGURE 6. Macro F_1 scores obtained by proposed models with different numbers of layers on SemEval-2018 Task-1C and GoEmotions datasets.

TABLE 7. F_1 Scores obtained with monolingual and multilingual embeddings.

	SemEval-2018 Task-1C dataset		GoEmotions Dataset	
PLM	UCCA-GAT	Dep-GAT	UCCA-GAT	Dep-GAT
	Monolingual Embeddings			
BERT	58.0	56.3	61.5	58.9
RoBERTa	59.2	56.9	**62.7**	59.5
XLNet	56.5	54.0	59.2	57.0
	Multilingual Embeddings			
M-BERT	58.5	57.8	63.5	61.1
XLM-R	60.0	56.2	**63.9**	60.2

(Average no. of words: SemEval-2018 Task-1C = 16.06, GoEmotions = 12.84) [103].

- The SemEval-2018 Task-1C dataset comprised of Tweets containing hashtags, emotions (labels in text), and punctuation marks, while the GoEmotions dataset comprised Reddit comments containing fewer emotion words, and no hashtags as compared to SemEval-2018.

However, some limitations of this study are dependency on external parser models, the difference between domains used to train parser models and the multi-label classification problem, and finally, emotions and punctuation marks in the datasets.

2) EMBEDDINGS

The effect of embeddings on models' success is huge [116]. Therefore, we tried monolingual and multilingual pre-trained embeddings in our experimentation to understand the behavior of the semantically and syntactically aware models. We used BERT [83], RoBERTa [107], and XLNet [117] monolingual embeddings with base variants consisting of 768 hidden dimensions, whereas we used multilingual version of BERT (M-BERT) [83] and RoBERTa (XLM-R) [118].

The results obtained by monolingual and multilingual pre-trained embeddings are in Table 7. The results show that multilingual embeddings are more effective for both proposed UCCA-GAT and Dep-GAT models.

3) IMPACT OF LAYERS

To further analyze the layers' impact on the proposed model, we varied the number of layers from 1 to 6 and performed experiments using BERT multilingual embeddings. Figure 6 illustrated the obtained results for the SemEval-2018 Task-1C and GoEmotions datasets.

As shown in Figure 6a, we achieved the highest results on the 3rd layer of the UCCA-GAT and Dep-GAT models for the SemEval-2018 Task-1C dataset. After the 3rd layer, there are a dramatic drop in the F_1 scores of the proposed models. This shows that deeper models lose the semantic features needed for the task. Figure 6b displayed the results of the GoEmotions dataset, where the highest performance can be seen on the 2nd layer of UCCA-GAT and 1st layer of Dep-GAT models. These results are similar to the SemEval-2018 Task-1C dataset. However, the dramatic drop can be seen in F_1 scores of both proposed models (UCCA-GAT and Dep-GAT) after the 4th layer.

The results obtained for the two datasets show that the semantically and syntactically aware models (UCCA-GAT and Dep-GAT) do not need deeper layers for the multi-label classification problem.

V. CONCLUSION AND FUTURE WORK

In recent years, semantically and syntactically aware models have gained popularity due to their impressive performance in NLP problems. However, these models have not been explored for the multi-label emotion classification problem. As described in this paper, our novel contribution is to develop a semantically and syntactically aware graph attention network for multi-label emotion classification problems in English texts using a challenging Twitter dataset provided by the SemEval-2018 E-c shared task and the GoEmotions dataset. We proposed a graph attention network using semantic and syntactic structures as input to the model. Our two proposed semantically and syntactically aware UCCA-GAT (accuracy = 71.2) and Dep-GAT (accuracy = 68.7) models

TABLE 8. Percentage of texts that were annotated with a given emotion in the datasets.

SemEval-2018 Task-1C		GoEmotions		GoEmotions (mapped to Ekman emotions)	
Emotion	**%**	**Emotion**	**%**	**Emotion**	**%**
Anger	36.1	Admiration	9.51	Anger	12.85
Anticipation	13.9	Amusement	5.36	Disgust	1.83
Disgust	36.6	Anger	3.61	Fear	1.67
Fear	16.8	Annoyance	5.69	Joy	40.11
Joy	39.3	Approval	6.77	Sadness	7.52
Love	12.3	Caring	2.50	Surprise	12.36
Optimism	31.3	Confusion	3.15	Neutral	32.76
Pessimism	11.6	Curiosity	5.05	-	-
Sadness	29.4	Desire	1.48	-	-
Surprise	5.2	Disappointment	2.92	-	-
Trust	5.0	Disapproval	4.66	-	-
Neutral	2.7	Disgust	1.83	-	-
-	-	Embarrassment	0.70	-	-
-	-	Excitement	1.96	-	-
-	-	Fear	1.37	-	-
-	-	Gratitude	6.13	-	-
-	-	Grief	0.18	-	-
-	-	Joy	3.34	-	-
-	-	Love	4.80	-	-
-	-	Nervousness	0.38	-	-
-	-	Optimism	3.64	-	-
-	-	Pride	0.26	-	-
-	-	Realization	2.56	-	-
-	-	Relief	0.35	-	-
-	-	Remorse	1.26	-	-
-	-	Sadness	3.05	-	-
-	-	Surprise	2.44	-	-
-	-	Neutral	32.76	-	-

outperformed the state-of-the-art studies on both datasets, i.e., GoEmotions and SemEval-2018.

In the future, we plan to examine other graph-based semantic representations, such as Abstract Meaning Representation, Prague Semantic Dependencies, and Elementary Dependency Structures, and investigate different models suitable for semantic graphs. Although the domains of the datasets are different, it would be beneficial to transfer learning between datasets, especially for the comparatively smaller SemEval-2018 Task-1C dataset. Therefore, after consulting with an expert to map SemEval-2018 Task-1C dataset emotions into Ekman's category, we plan to extend our experiments with transfer learning. We also plan to further improvements in samplings to address the problem of imbalance dataset for multi-label emotion classification.

APPENDIX I. DETAILS OF DATASETS

Table 8 indicates the percentage of texts annotated with given emotions in the datasets. We also added the GoEmotions dataset before mapping the labels into Ekman's categories to show how imbalanced the original dataset is. The rows sum to more than 100% because a text is likely to contain more than one emotion. Note that anger, disgust, joy, optimism, and sadness received a higher percentage of emotion labels, while neutral, surprise, and trust are the least annotated emotions in the SemEval-2018 Task-1C dataset. Similarly, joy and neutral received a higher percentage in the GoEmotions dataset.

APPENDIX II. HYPERPARAMETER VALUES

Table 9 lists the hyperparameter values used in the model.

TABLE 9. Hyperparameters used for the different models in experiments.

Parameters	SemEval-2018		GoEmotions	
	UCCA-GAT	Dep-GAT	UCCA-GAT	Dep-GAT
weight decay	0.1	0.2	0.2	0.2
batch size	1	1	1	1
learning rate	0.001	0.005	0.001	0.005
dropout rate	0.2	0.1	0.1	0.2
number of head	2	4	2	4

ACKNOWLEDGMENT

(Iqra Ameer and Necva Bölücü contributed equally to this work.) The authors thank the CONACYT for the computing resources brought to them through Plataforma de Aprendizaje Profundo para Tecnologías del Lenguaje of the Laboratorio de Supercómputo of the INAOE, Mexico. They also acknowledge the support of Microsoft through the Microsoft Latin America Ph.D. Award.

REFERENCES

[1] H. C. Hwang and D. Matsumoto, "Facial expressions," in *APA Handbook of Nonverbal Communication*. Washington, DC, USA: American Psychological Association, 2016, pp. 257–287.

[2] J. Kumar and J. Kumar, "Machine learning approach to classify emotions using GSR," *Adv. Res. Electr. Electron. Eng.*, vol. 2, no. 12, pp. 72–76, 2015.

[3] A. R. Damasio, T. J. Grabowski, A. Bechara, H. Damasio, L. L. B. Ponto, J. Parvizi, and R. D. Hichwa, "Subcortical and cortical brain activity during the feeling of self-generated emotions," *Nature Neurosci.*, vol. 3, no. 10, pp. 1049–1056, Oct. 2000.

[4] P. Ekman and H. Oster, "Facial expressions of emotion," *Annu. Rev. Psychol.*, vol. 30, no. 1, pp. 527–554, 1979.

[5] R. Plutchik, "A general psychoevolutionary theory of emotion," in *Theories of Emotion*. Amsterdam, The Netherlands: Elsevier, 1980, pp. 3–33.

[6] W. G. Parrott, *Emotions in Social Psychology: Essential Readings*. London, U.K.: Psychology Press, 2001.

[7] V. Revathy and A. S. Pillai, "Binary emotion classification of music using deep neural networks," in *Proc. 13th Int. Conf. Soft Comput. Pattern Recognit.* Cham, Switzerland: Springer, 2022, pp. 484–492.

[8] R. C. Balabantaray, M. Mohammad, and N. Sharma, "Multi-class Twitter emotion classification: A new approach," *Int. J. Appl. Inf. Syst.*, vol. 4, no. 1, pp. 48–53, Sep. 2012.

[9] S. Chawla and M. Mehrotra, "An ensemble-classifier based approach for multiclass emotion classification of short text," in *Proc. 7th Int. Conf. Rel., INFOCOM Technol. Optim. (Trends Future Directions) (ICRITO)*, Aug. 2018, pp. 768–774.

[10] S. Mohammad, F. Bravo-Marquez, M. Salameh, and S. Kiritchenko, "SemEval-2018 task 1: Affect in tweets," in *Proc. 12th Int. Workshop Semantic Eval.*, New Orleans, LA, USA, 2018, pp. 1–17.

[11] I. Ameer, G. Sidorov, H. Gómez-Adorno, and R. M. A. Nawab, "Multi-label emotion classification on code-mixed text: Data and methods," *IEEE Access*, vol. 10, pp. 8779–8789, 2022.

[12] S. Zhang, Y. Li, Y. Liu, and H. Chen, "Emotional advertising: A study of the emotional impact of advertising on consumer behavior," *J. Financial Services Marketing*, vol. 21, no. 4, pp. 288–299, 2016.

[13] Y. Yang, Y. Liu, and J. Mao, "Sentiment analysis of financial news and its impact on stock price movements," *J. Finance Data Sci.*, vol. 6, no. 3, pp. 310–320, 2020.

[14] E. I. Lysova and E. I. Rasskazova, "Emotions in financial decision-making: A systematic review," *J. Behav. Experim. Finance*, vol. 24, pp. 100–113, Jan. 2019.

[15] K. Lin, F. Xia, W. Wang, D. Tian, and J. Song, "System design for big data application in emotion-aware healthcare," *IEEE Access*, vol. 4, pp. 6901–6909, 2016.

[16] A. H. Saffar, T. K. Mann, and B. Ofoghi, "Textual emotion detection in health: Advances and applications," *J. Biomed. Informat.*, vol. 137, Jan. 2023, Art. no. 104258.

[17] C.-W. Huang, B. C. Y. Wu, P. A. Nguyen, H.-H. Wang, C.-C. Kao, P.-C. Lee, A. R. Rahmanti, J. C. Hsu, H.-C. Yang, and Y.-C.-J. Li, "Emotion recognition in doctor-patient interactions from real-world clinical video database: Initial development of artificial empathy," *Comput. Methods Programs Biomed.*, vol. 233, May 2023, Art. no. 107480.

[18] Y. Huang and J. Zhang, "A study of the effectiveness of emotion recognition for student learning outcomes in e-learning environments," *Interact. Learn. Environ.*, vol. 27, no. 7, pp. 1019–1032, 2019.
[19] L. Zhang, J. Yang, and Z. Tang, "Designing emotional intelligent tutors: A systematic review of affective computing in education," *J. Educ. Comput. Res.*, vol. 57, no. 8, pp. 2133–2166, 2020.
[20] A. Carstens, V. Hockly, M. P. Koen, and E. J. Pretorius, "An investigation into the use of emotional intelligence for learning analytics," *Int. J. Educ. Technol. Higher Educ.*, vol. 16, no. 1, pp. 1–15, 2019.
[21] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi, "GoEmotions: A dataset of fine-grained emotions," 2020, *arXiv:2005.00547*.
[22] M. A. Mahima, N. C. Patel, S. Ravichandran, N. Aishwarya, and S. Maradithaya, "A text-based hybrid approach for multiple emotion detection using contextual and semantic analysis," in *Proc. Int. Conf. Innov. Comput., Intell. Commun. Smart Electr. Syst. (ICSES)*, Sep. 2021, pp. 1–6.
[23] R. Kamath, A. Ghoshal, S. Eswaran, and P. B. Honnavalli, "EmoRoBERTa: An enhanced emotion detection model using RoBERTa," in *Proc. IEEE Int. Conf. Electron., Comput. Commun. Technol.*, Apr. 2022, pp. 1–6.
[24] A. Kane, S. Patankar, S. Khose, and N. Kirtane, "Transformer based ensemble for emotion detection," 2022, *arXiv:2203.11899*.
[25] I. Ameer, N. Ashraf, G. Sidorov, and H. G. Adorno, "Multi-label emotion classification using content-based features in Twitter," *Computación Y Sistemas*, vol. 24, no. 3, pp. 1159–1164, Sep. 2020.
[26] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
[27] A. Graves, "Long short-term memory," in *Supervised Sequence Labelling with Recurrent Neural Networks*. Berlin, Germany: Springer, 2012, pp. 37–45.
[28] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997.
[29] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," 2014, *arXiv:1406.1078*.
[30] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Rev. Neurosci.*, vol. 2, no. 3, pp. 194–203, Mar. 2001.
[31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
[32] Y. Lv, F. Wei, L. Cao, S. Peng, J. Niu, S. Yu, and C. Wang, "Aspect-level sentiment analysis using context and aspect memory network," *Neurocomputing*, vol. 428, pp. 195–205, Mar. 2021.
[33] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, Jan. 2020.
[34] T. Ji, Y. Wu, and M. Lan, "Graph-based dependency parsing with graph neural networks," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 2475–2485.
[35] H. Yan, X. Qiu, and X. Huang, "A graph-based model for joint Chinese word segmentation and dependency parsing," *Trans. Assoc. Comput. Linguistics*, vol. 8, pp. 78–92, Dec. 2020.
[36] A. Hamid, N. Shiekh, N. Said, K. Ahmad, A. Gul, L. Hassan, and A. Al-Fuqaha, "Fake news detection in social media using graph neural networks and NLP techniques: A COVID-19 use-case," 2020, *arXiv:2012.07517*.
[37] M. Malekzadeh, P. Hajibabaee, M. Heidari, S. Zad, O. Uzuner, and J. H. Jones, "Review of graph neural network in text classification," in *Proc. IEEE 12th Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf. (UEMCON)*, Dec. 2021, pp. 0084–0091.
[38] O. Abend and A. Rappoport, "The state of the art in semantic representation," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics*, 2017, pp. 77–89.
[39] L. Banarescu, C. Bonial, S. Cai, M. Georgescu, K. Griffitt, U. Hermjakob, K. Knight, P. Koehn, M. Palmer, and N. Schneider, "Abstract meaning representation for sembanking," in *Proc. 7th Linguistic Annotation Workshop Interoperability Discourse*, 2013, pp. 178–186.
[40] O. Abend and A. Rappoport, "UCCA: A semantics-based grammatical annotation scheme," in *Proc. 10th Int. Conf. Comput. Semantics*, 2013, pp. 1–12.
[41] S. Oepen, M. Kuhlmann, Y. Miyao, D. Zeman, S. Cinkova, D. Flickinger, J. Hajic, A. Ivanova, and Z. Uresova, "Towards comparability of linguistic graph banks for semantic parsing," in *Proc. 10th Int. Conf. Lang. Resour. Eval.*, 2016, pp. 3991–3995.
[42] A. S. White, D. Reisinger, K. Sakaguchi, T. Vieira, S. Zhang, R. Rudinger, K. Rawlins, and B. Van Durme, "Universal decompositional semantics on universal dependencies," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 1713–1723.
[43] L. Abzianidze, J. Bjerva, K. Evang, H. Haagsma, R. van Noord, P. Ludmann, D.-D. Nguyen, and J. Bos, "The parallel meaning bank: Towards a multilingual corpus of translations annotated with compositional meaning representations," in *Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguistics*, 2017, pp. 242–247.
[44] S. Dohare, H. Karnick, and V. Gupta, "Text summarization using abstract meaning representation," 2017, *arXiv:1706.01678*.
[45] F. Liu, J. Flanigan, S. Thomson, N. Sadeh, and N. A. Smith, "Toward abstractive summarization using semantic representations," 2018, *arXiv:1805.10399*.
[46] F. Issa, M. Damonte, S. B. Cohen, X. Yan, and Y. Chang, "Abstract meaning representation for paraphrase detection," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Human Lang. Technol.*, 2018, pp. 442–452.
[47] R. Blloshmi, R. Tripodi, and R. Navigli, "XL-AMR: Enabling cross-lingual AMR parsing with transfer learning techniques," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2020, pp. 2487–2500.
[48] L. Song, D. Gildea, Y. Zhang, Z. Wang, and J. Su, "Semantic neural machine translation using AMR," *Trans. Assoc. Comput. Linguistics*, vol. 7, pp. 19–31, Mar. 2019.
[49] E. Sulem, O. Abend, and A. Rappoport, "Semantic structural decomposition for neural machine translation," in *Proc. 9th Joint Conf. Lexical Comput. Semantics*, 2020, pp. 50–57.
[50] P. Kapanipathi, I. Abdelaziz, S. Ravishankar, S. Roukos, A. Gray, R. Astudillo, M. Chang, C. Cornelio, S. Dana, A. Fokoue, and D. Garg, "Leveraging abstract meaning representation for knowledge base question answering," 2020, *arXiv:2012.01707*.
[51] W. Xu, H. Zhang, D. Cai, and W. Lam, "Dynamic semantic graph construction and reasoning for explainable multi-hop science question answering," 2021, *arXiv:2105.11776*.
[52] E. Sulem, O. Abend, and A. Rappoport, "Simple and effective text simplification using semantic and neural methods," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, 2018, pp. 162–173.
[53] H. Ren, Y. Ren, X. Li, W. Feng, and M. Liu, "Natural logic inference for emotion detection," in *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*. Berlin, Germany: Springer, 2017, pp. 424–436.
[54] D. Tang, Z. Zhang, Y. He, C. Lin, and D. Zhou, "Hidden topic-emotion transition model for multi-level social emotion detection," *Knowl.-Based Syst.*, vol. 164, pp. 426–435, Jan. 2019.
[55] D. Zhang, X. Ju, J. Li, S. Li, Q. Zhu, and G. Zhou, "Multi-modal multi-label emotion detection with modality and label dependence," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2020, pp. 3584–3593.
[56] R. Wang, A. Finch, M. Utiyama, and E. Sumita, "Sentence embedding for neural machine translation domain adaptation," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics*, 2017, pp. 560–566.
[57] X. Wu, C. Gao, L. Zang, J. Han, Z. Wang, and S. Hu, "ESimCSE: Enhanced sample building method for contrastive learning of unsupervised sentence embedding," 2021, *arXiv:2109.04380*.
[58] J. Zhang and Z. Lan, "S-SimCSE: Sampled sub-networks for contrastive learning of sentence embedding," 2021, *arXiv:2111.11750*.
[59] C. Baziotis, N. Athanasiou, A. Chronopoulou, A. Kolovou, G. Paraskevopoulos, N. Ellinas, S. Narayanan, and A. Potamianos, "NTUA-SLP at SemEval-2018 task 1: Predicting affective content in tweets with deep attentive RNNs and transfer learning," 2018, *arXiv:1804.06658*.
[60] G. Gee and E. Wang, "PsyML at SemEval-2018 task 1: Transfer learning for sentiment and emotion analysis," in *Proc. 12th Int. Workshop Semantic Eval.*, 2018, pp. 369–376.

[61] Y. Kim, H. Lee, and K. Jung, "AttnConvnet at SemEval-2018 task 1: Attention-based convolutional neural networks for multi-label emotion classification," 2018, *arXiv:1804.00831*.

[62] B. Eisner, T. Rocktäschel, I. Augenstein, M. Bosnjak, and S. Riedel, "emoji2vec: Learning emoji representations from their description," 2016, *arXiv:1609.08359*.

[63] N. Alswaidan and M. E. B. Menai, "A survey of state-of-the-art approaches for emotion recognition in text," *Knowl. Inf. Syst.*, vol. 62, pp. 2937–2987, Aug. 2020.

[64] F. Ding, X. Kang, S. Nishide, Z. Guan, and F. Ren, "A fusion model for multi-label emotion classification based on BERT and topic clustering," in *Proc. Int. Symp. Artif. Intell. Robot.*, Oct. 2020, pp. 98–111.

[65] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 3111–3119.

[66] H. Meisheri and L. Dey, "TCS research at SemEval-2018 task 1: Learning robust representations using multi-attention architecture," in *Proc. 12th Int. Workshop Semantic Eval.*, 2018, pp. 291–299.

[67] A. Radford, R. Jozefowicz, and I. Sutskever, "Learning to generate reviews and discovering sentiment," 2017, *arXiv:1704.01444*.

[68] S. Rajalakshmi, S. M. Rajendram, and T. Mirnalinee, "SSN MLRG1 at SemEval-2018 task 1: Emotion and sentiment intensity detection using rule based feature selection," in *Proc. 12th Int. Workshop Semantic Eval.*, 2018, pp. 324–328.

[69] P. Rathnayaka, S. Abeysinghe, C. Samarajeewa, I. Manchanayake, M. J. Walpola, R. Nawaratne, T. Bandaragoda, and D. Alahakoon, "Gated recurrent neural network approach for multilabel emotion detection in microblogs," 2019, *arXiv:1907.07653*.

[70] S. Park, J. Kim, S. Ye, J. Jeon, H. Y. Park, and A. Oh, "Dimensional emotion detection from categorical emotion," 2019, *arXiv:1911.02499*.

[71] I. Ameer, N. Bölucu, M. H. F. Siddiqui, B. Can, G. Sidorov, and A. Gelbukh, "Multi-label emotion classification in texts using transfer learning," *Exp. Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 118534.

[72] J. Li and L. Xiao, "Multi-emotion recognition using multi-EmoBERT and emotion analysis in fake news," in *Proc. 15th ACM Web Sci. Conf.*, New York, NY, USA, Apr. 2023, pp. 128–135, doi: 10.1145/3578503.3583595.

[73] N. Farruque, C. Huang, O. Zaiane, and R. Goebel, "Basic and depression specific emotions identification in tweets: Multi-label classification experiments," in *Proc. Int. Conf. Comput. Linguistics Intell. Text Process.* La Rochelle, France: Springer, 2023, pp. 293–306.

[74] S. Islam, A. C. Roy, M. S. Arefin, and S. Afroz, "Multi-label emotion classification of tweets using machine learning," in *Proc. Int. Conf. Big Data, IoT, Mach. Learn.* Cham, Switzerland: Springer, 2022, pp. 705–722.

[75] U. Ahmed, L. Zafar, F. Qayyum, and M. A. Islam, "Irony detector at SemEval-2018 task 3: Irony detection in English tweets using word graph," in *Proc. 12th Int. Workshop Semantic Eval.*, 2018, pp. 581–586.

[76] E. Elbasani and J.-D. Kim, "AMR-CNN: Abstract meaning representation with convolution neural network for toxic content detection," *J. Web Eng.*, vol. 21_3, pp. 677–692, Feb. 2022.

[77] Y. Bao, Q. Ma, L. Wei, W. Zhou, and S. Hu, "Multi-granularity semantic aware graph model for reducing position bias in emotion-cause pair extraction," 2022, *arXiv:2205.02132*.

[78] B. Galitsky, "Employing abstract meaning representation to lay the last-mile toward reading cmprehension," in *Artificial Intelligence for Customer Relationship Management*. Berlin, Germany: Springer, 2020, pp. 57–86.

[79] Y. Qin and Y. Liang, "Semantic analysis and evaluation of translation based on abstract meaning representation," in *Proc. Int. Conf. Web Inf. Syst. Appl.* Cham, Switzerland: Springer, 2020, pp. 268–275.

[80] J. Xu, Y. Guo, and J. Hu, "Incorporate semantic structures into machine translation evaluation via UCCA," 2020, *arXiv:2010.08728*.

[81] Z. Zhang, Y. Wu, H. Zhao, Z. Li, S. Zhang, X. Zhou, and X. Zhou, "Semantics-aware BERT for language understanding," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 5, pp. 9628–9635.

[82] M. Palmer, D. Gildea, and P. Kingsbury, "The proposition bank: An annotated corpus of semantic roles," *Comput. Linguistics*, vol. 31, no. 1, pp. 71–106, Mar. 2005.

[83] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

[84] A. Ek, J.-P. Bernardy, and S. Lappin, "Language modeling with syntactic and semantic representation for sentence acceptability predictions," in *Proc. 22nd Nordic Conf. Comput. Linguistics*, 2019, pp. 76–85.

[85] J. Nivre, M.-C. de Marneffe, F. Ginter, J. Hajic, C. D. Manning, S. Pyysalo, S. Schuster, F. Tyers, and D. Zeman, "Universal dependencies v2: An evergrowing multilingual treebank collection," 2020, *arXiv:2004.10643*.

[86] L. H. B. Nguyen, V. H. Pham, and D. Dinh, "Improving neural machine translation with AMR semantic graphs," *Math. Problems Eng.*, vol. 2021, pp. 1–12, Jul. 2021.

[87] A. Slobodkin, L. Choshen, and O. Abend, "Semantics-aware attention improves neural machine translation," 2021, *arXiv:2110.06920*.

[88] L. Huang, X. Sun, S. Li, L. Zhang, and H. Wang, "Syntax-aware graph attention network for aspect-level sentiment classification," in *Proc. 28th Int. Conf. Comput. Linguistics*, 2020, pp. 799–810.

[89] Q. Guo, X. Qiu, X. Xue, and Z. Zhang, "Syntax-guided text generation via graph neural network," *Sci. China Inf. Sci.*, vol. 64, no. 5, pp. 1–10, May 2021.

[90] M. S. Schlichtkrull, N. De Cao, and I. Titov, "Interpreting graph neural networks for NLP with differentiable edge masking," 2020, *arXiv:2010.00577*.

[91] J. Bastings, I. Titov, W. Aziz, D. Marcheggiani, and K. Sima'an, "Graph convolutional encoders for syntax-aware neural machine translation," 2017, *arXiv:1704.04675*.

[92] A. Nagy, P. Nanys, B. F. Konrád, B. Bial, and J. Ács, "Syntax-based data augmentation for Hungarian–English machine translation," 2022, *arXiv:2201.06876*.

[93] Z. Li, Q. Zhou, C. Li, K. Xu, and Y. Cao, "Improving BERT with syntax-aware local attention," 2020, *arXiv:2012.15150*.

[94] J. Bai, Y. Wang, Y. Chen, Y. Yang, J. Bai, J. Yu, and Y. Tong, "Syntax-BERT: Improving pre-trained transformers with syntax trees," 2021, *arXiv:2103.04350*.

[95] D. Marcheggiani and I. Titov, "Graph convolutions over constituent trees for syntax-aware semantic role labeling," 2019, *arXiv:1909.09814*.

[96] X.-P. Nguyen, S. Joty, S. C. H. Hoi, and R. Socher, "Tree-structured attention with hierarchical accumulation," 2020, *arXiv:2002.08046*.

[97] X. Carreras and L. Marquez, "Introduction to the CoNLL-2005 shared task: Semantic role labeling," in *Proc. 9th Conf. Comput. Natural Lang. Learn.*, Ann Arbor, Michigan, 2005, pp. 152–164. [Online]. Available: https://aclanthology.org/W05-0620

[98] S. Pradhan, A. Moschitti, N. Xue, O. Uryupina, and Y. Zhang, "CoNLL-2012 shared task: Modeling multilingual unrestricted coreference in OntoNotes," in *Proc. Joint Conf. EMNLP CoNLL-Shared Task*, Jeju Island, (South) Korea, Jul. 2012, pp. 1–40. [Online]. Available: https://aclanthology.org/W12-4501

[99] C. F. Baker, C. J. Fillmore, and J. B. Lowe, "The Berkeley FrameNet project," in *Proc. 17th Int. Conf. Comput. Linguistics*, 1998, pp. 86–90.

[100] L. Cagliero and P. Garza, "Improving classification models with taxonomy information," *Data Knowl. Eng.*, vol. 86, pp. 85–101, Jul. 2013.

[101] M. T. Pilehvar and J. Camacho-Collados, "Embeddings in natural language processing: Theory and advances in vector representations of meaning," *Synth. Lectures Human Lang. Technol.*, vol. 13, no. 4, pp. 1–175, Nov. 2020.

[102] Q. P. Zeng and S. X. Wu, "Using semantic constraints for question answering," in *Advanced Materials Research*, vol. 655. Zürich, Switzerland: Trans Tech Publications Ltd, 2013, pp. 1750–1756.

[103] N. Bölucu and B. Can, "Self-attentive constituency parsing for UCCA-based semantic parsing," 2021, *arXiv:2110.00621*.

[104] J.-C. Chappelier and M. Rajman, "A generalized CYK algorithm for parsing stochastic CFG," in *Proc. 1st Workshop Tabulation Parsing Deductio*, 1998, pp. 133–137.

[105] C. Arora, M. Sabetzadeh, L. Briand, and F. Zimmer, "Extracting domain models from natural-language requirements: Approach and industrial evaluation," in *Proc. ACM/IEEE 19th Int. Conf. Model Driven Eng. Lang. Syst.*, Oct. 2016, pp. 250–260.

[106] T. Dozat and C. D. Manning, "Deep biaffine attention for neural dependency parsing," 2016, *arXiv:1611.01734*.

[107] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," 2019, *arXiv:1907.11692*.

[108] S. van der Walt, S. C. Colbert, and G. Varoquaux, "The NumPy array: A structure for efficient numerical computation," *Comput. Sci. Eng.*, vol. 13, no. 2, pp. 22–30, Mar. 2011.

[109] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*.
[110] H. Cai, S. Lv, G. Lu, and T. Li, "Graph convolutional networks for fast text classification," in *Proc. 4th Int. Conf. Natural Lang. Process. (ICNLP)*, Mar. 2022, pp. 420–425.
[111] D. Hershcovich, Z. Aizenbud, L. Choshen, E. Sulem, A. Rappoport, and O. Abend, "SemEval-2019 task 1: Cross-lingual semantic parsing with UCCA," in *Proc. 13th Int. Workshop Semantic Eval.*, 2019, pp. 1–10.
[112] P. Qi, Y. Zhang, Y. Zhang, J. Bolton, and C. D. Manning, "Stanza: A Python natural language processing toolkit for many human languages," 2020, *arXiv:2003.07082*.
[113] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, and L. Antiga, "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–12.
[114] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[115] J. Ho Park, P. Xu, and P. Fung, "PlusEmo2Vec at SemEval-2018 task 1: Exploiting emotion knowledge from emoji and #hashtags," 2018, *arXiv:1804.08280*.
[116] L. Zhang, T. Xiang, and S. Gong, "Learning a deep embedding model for zero-shot learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3010–3019.
[117] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized autoregressive pretraining for language understanding," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–11.
[118] A. Conneau and G. Lample, "Cross-lingual language model pretraining," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–10.

IQRA AMEER received the B.S. degree in computer science from COMSATS University Islamabad, Lahore, Pakistan, and the M.S. and Ph.D. degrees (Hons.) in computer science from Centro de Investigación en Computación, Instituto Politécnico Nacional, Mexico City, Mexico, in 2017 and 2022, respectively. She is currently an Assistant Professor with the Division of Engineering and Science, Abington College, Pennsylvania State University, PA, USA. Previously, she was a Postdoctoral Research Associate with Yale University. Her current research interests include natural language processing and the application of machine learning techniques to NLP tasks.

NECVA BÖLÜCÜ received the M.Sc. and Ph.D. degrees in computer engineering from Hacettepe University, Turkey, in 2017 and 2022, respectively. She is a Postdoctoral Fellow with CSIRO. From 2015 to 2022, she was a Research Assistant with Hacettepe University and Alparslan Turkes Science and Technology University. Her research focuses on semantic graph representation from a computational perspective. Her research is linked to the evaluation of meaning representations in NLP problems, such as semantic textual similarity and text classification.

GRIGORI SIDOROV received the Ph.D. degree in philology (computational and structural linguistics) from Lomonosov Moscow State University, Russia, in 1996. He is currently a Full Professor with the Natural Language Processing Laboratory, Centro de Investigación en Computación, Instituto Politécnico Nacional, Mexico City, Mexico. He is a National Researcher of Mexico Level 3 (highest level, about 2-3% of all researchers in Mexico have this level). He is the author of more than 350 scientific publications, of which nine books. His current research interests include computational linguistics, natural language processing, and the application of machine learning techniques to NLP tasks. He is a member of the Mexican Academy of Sciences. He is the Editor-in-Chief of the research journal *Computación y Sistemas* (WoS, Scopus, and CONACYT).

BURCU CAN received the B.S. and M.Sc. degrees in computer engineering from Hacettepe University, Turkey, in 2005 and 2007, respectively, and the Ph.D. degree in computer science from the University of York, U.K., in 2011. She was a Visiting Scholar with the Artificial Intelligence Group, University of York, from 2014 to 2015, and a Visiting Researcher with the Institute of Statistical Mathematics, Tokyo, Japan, in 2019. She is currently a Lecturer with the Department of Computing Science and Mathematics, University of Stirling. Her current research interests include natural language processing and a range of machine learning models, such as Bayesian learning and deep neural networks.