Article

Syntactic Dependency-Aware Neural Networks for Enhanced Entity Relation Analysis

Muhammad Nazrul , Carlos Stellato , Sarker Iqbal , Fei Lee * , Woods Ali

Posted Date: 27 November 2023

doi: 10.20944/preprints202311.1681.v1

Keywords: information extraction; entity relation extraction; graph neural networks; dependency feature

Article

Syntactic Dependency-Aware Neural Networks for Enhanced Entity Relation Analysis

Muhammad Nazrul [1], Carlos Stellato [1], Sarker Iqbal [1], Fei Lee [2,*] and Woods Ali [1]

[1] University of Charleston
[2] National University of Singapore
[*] Correspondence: feih@comp.nus.edu.sg

Abstract: In the intricate domain of text analysis, syntactic dependency trees play a crucial role in unraveling the web of relations among entities embedded in the text. These trees provide a structural roadmap, guiding us through the complex syntax to pinpoint the interactions and connections between different entities. However, the challenge lies in sifting through this intricate structure to extract relevant information, a task that requires precision and discernment. Traditional approaches often rely on rule-based pruning methods to simplify these dependency structures, focusing on certain parts while discarding others. Yet, this approach has its pitfalls, as it can overlook critical nuances and connections that are vital for a comprehensive understanding of the text. Addressing this gap, our research introduces the Syntactic Dependency-Aware Neural Networks (SDANNs), a groundbreaking model designed to harness the full power of the entire dependency tree. This approach marks a significant departure from the conventional methods. Instead of the rigid rule-based pruning, SDANNs implement a more flexible and dynamic 'soft-pruning' technique. This method allows the model to adaptively focus on the sub-structures within the dependency tree that are most relevant for understanding the relationships between entities. By doing so, it ensures that no vital information is overlooked, and all potential connections are considered. The efficacy of SDANNs is not just theoretical but has been empirically validated through extensive testing and evaluations across a wide range of tasks. These tasks include the extraction of complex relations spanning multiple sentences, as well as detailed analyses at the sentence level. In each of these scenarios, SDANNs have demonstrated a remarkable ability to leverage the full structural complexity of dependency trees. This capability sets them apart from existing models, enabling a more nuanced and comprehensive analysis of textual relations. The results of these evaluations consistently show that SDANNs not only meet but significantly exceed the performance of prior models. This superiority is evident in the way SDANNs handle the multifaceted and often subtle interactions within the text, offering insights that were previously inaccessible with conventional methods. In summary, the Syntactic Dependency-Aware Neural Networks represent a significant advancement in the field of text analysis. By fully embracing the complexity of syntactic dependency trees and employing a sophisticated 'soft-pruning' approach, SDANNs open new avenues for exploring and understanding the intricate relationships that exist within written language. This model stands as a testament to the potential of combining advanced neural network architectures with a deep understanding of linguistic structures, paving the way for more accurate, nuanced, and comprehensive analyses of text.

Keywords: information extraction; entity relation extraction; graph neural networks; dependency feature

1. Introduction

Identifying relations among textual entities is crucial for numerous language processing applications, such as biomedical data mining (Quirk and Poon 2017), database enrichment (Zhang *et al.* 2017), and interactive query systems (Yu *et al.* 2017). Consider, for instance, a scenario detailing a relationship among the entities *L858E*, *EGFR*, and *gefitinib* across two sentences.

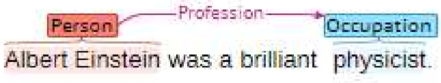

Figure 1. An example for entity relation extraction.

Relation extraction techniques primarily divide into two categories: sequence-oriented and dependency-focused. The former relies solely on word sequences (Zeng *et al.* 2014; Wang *et al.* 2016), while the latter incorporates syntactic dependency trees (Bunescu and Mooney 2005; Peng *et al.* 2017). Dependency-focused methods surpass their sequence-oriented counterparts by capturing syntactic relations that are not immediately evident. Various pruning techniques have been suggested to refine dependency data, thereby enhancing performance. For instance, Xu *et al.* (2015); Fei *et al.* (2022); Xu *et al.* (2015) employ neural networks on the shortest dependency paths, while Miwa and Bansal (2016) concentrate on subtrees linked to the lowest common ancestor (LCA) of the entities. Zhang *et al.* (2018) apply graph convolutional networks (GCNs) (Kipf and Welling 2017) to pruned trees, including nodes within a certain distance from the dependency path in the LCA subtree.

However, these rule-based pruning approaches risk losing crucial data from the complete tree. For example, critical tokens like *partial response* might be overlooked if only pruned trees are considered. Ideally, models should adeptly include or exclude data from the full tree (Tang *et al.* 2016; Li *et al.* 2022; Fei *et al.* 2021; Dong *et al.* 2014; Fei *et al.* 2020). Our paper introduces the SDANNs, functioning directly on the full tree and employing a 'soft pruning' approach. This method transforms the dependency tree into a fully-connected, edge-weighted graph, with self-attention mechanisms (Vaswani *et al.* 2017) determining the relevance of connections.

To effectively process these dense graphs, we integrate dense connections into the GCN model as per (Huang *et al.* 2017; Guo *et al.* 2019). Standard GCNs require multiple layers to capture multi-hop neighborhood information, yet deeper models often do not yield proportional benefits. With dense connections, our SDANN model can be deeply layered, effectively capturing both local and non-local dependency information.

Our experiments affirm the SDANN's enhanced performance in various scenarios. In multi-sentence relation extraction, our model excels, surpassing current leading models in both ternary and binary relation accuracy by significant margins. On the comprehensive sentence-level TACRED dataset, the SDANN consistently outperforms existing models. In summary, our proposed SDANNs introduce an end-to-end 'soft pruning' strategy for better graph representation learning, achieving state-of-the-art results without increased computational demands. The adjacency matrix size remains constant compared to the original tree, ensuring efficiency in parallel processing over dependency trees, unlike tree-structured models like Tree-LSTM (Tai *et al.* 2015).

2. Related Work

The foundation of our research is rooted in the evolving landscape of relation extraction models and the advancements in graph convolutional networks. In the early stages, relation extraction research predominantly hinged on statistical methodologies. Pioneers in this field explored tree-based (Zelenko *et al.* 2002) and dependency path-based kernels (Bunescu and Mooney 2005) to delve into the intricacies of relation extraction. Groundbreaking work by McDonald *et al.* (2005) involved constructing maximal cliques of entities, providing a novel perspective on predicting relations. Adding another dimension to this approach, Mintz *et al.* (2009) amalgamated syntactic features into statistical classifiers, enriching the analytical process.

With the advent of neural networks, the landscape shifted towards sequence-based models. These models harnessed the power of various neural network architectures for relation extraction. Convolutional neural networks (CNNs) were employed in groundbreaking studies (Zeng *et al.* 2014; Wang *et al.* 2016; Nguyen and Grishman 2015), while recurrent neural networks (RNNs) found their use in attention-based learning (Zhang *et al.* 2017; Zhou *et al.* 2016; Fei *et al.* 2023). The ingenuity of combining CNNs and RNNs was demonstrated by Vu *et al.* (2016), and the emergence of transformers (Verga *et al.* 2018) marked a significant evolution in this domain.

Dependency-based models also evolved, integrating structural information into neural frameworks. Peng *et al.* (2017) innovated by splitting the dependency graph into two Directed Acyclic Graphs (DAGs) and extending the tree LSTM model (Tai *et al.* 2015) for complex n-ary relation extraction. Aligning closely with our work, Song *et al.* (2018a) employed graph recurrent networks (Song *et al.* 2018b) to encode entire dependency graphs. This approach contrasts with ours similarly to the differences between CNNs and RNNs. Various pruning strategies were also proposed to distill dependency information for enhanced model performance. Techniques ranged from encoding the shortest dependency path (Xu *et al.* 2015?; Fei *et al.* 2023), applying LSTM models to the LCA subtree of entities (Miwa and Bansal 2016; Fei *et al.* 2020), combining dependency paths and subtrees (Liu *et al.* 2015), to adopting path-centric pruning strategies (Zhang *et al.* 2018). Our model diverges from these by learning to variably weigh each edge in an end-to-end manner, rather than removing edges in preprocessing.

The initial steps to adapt neural networks for arbitrary structured graphs were made by Gori *et al.* (2005); Bruna (2014). These foundational efforts paved the way for more computationally efficient adaptations through local spectral convolution techniques (Henaff *et al.* 2015; Defferrard *et al.* 2016). Our methodology closely aligns with the GCNs (Kipf and Welling 2017; Fei *et al.* 2020), which focus on filters operating within a first-order neighborhood around each node.

In a more recent and innovative stride, Velickovic *et al.* (2018) introduced graph attention networks (GATs), leveraging masked self-attentional layers (Vaswani *et al.* 2017) to effectively summarize neighborhood states. While sharing some conceptual similarities with our work, their motivations and network structures exhibit significant differences. In GATs, each node's focus is confined to its immediate neighbors, whereas our AGGCNs assess the relatedness across all nodes. This distinct approach allows for the construction of fully connected graphs in AGGCNs, enabling the capture of long-range semantic interactions, a feature not present in the static network topology of GATs.

3. Methodology

This section delineates the core elements constituting our Syntactic Dependency Graph Convolutional Network (SDANN) model.

3.1. Graph-Based Neural Networks

Graph-based Neural Networks (GNNs), a variant of neural networks tailored for graph data, operate directly on graphs (Kipf and Welling 2017). We mathematically demonstrate how GNNs function on a graph with n nodes, represented by an $n \times n$ adjacency matrix $\mathbf{A}$. Marcheggiani and Titov (2017) augmented GNNs to process dependency trees by integrating edge directionality. They introduce self-loops for each node, and the adjacency matrix reflects both directions of a dependency arc, i.e., $\mathbf{A}_{ij} = 1$ and $\mathbf{A}_{ji} = 1$ if there's an edge from node i to j, and zero otherwise. The convolution operation for node i at layer l, with input feature $\mathbf{h}^{(l-1)}$ and output $\mathbf{h}_i^{(l)}$, is defined as:

$$\mathbf{h}_i^{(l)} = \rho\Big(\sum_{j=1}^{n} \mathbf{A}_{ij}\mathbf{W}^{(l)}\mathbf{h}_j^{(l-1)} + \mathbf{b}^{(l)}\Big) \tag{1}$$

where $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$ denote the weight matrix and bias vector, respectively, and ρ is an activation function such as RELU. The initial input $\mathbf{h}_i^{(0)}$ is $\mathbf{x}_i$, with $\mathbf{x}_i \in \mathbb{R}^d$ and d being the input feature dimension.

3.2. Syntactic Attention Layer

The SDANN comprises M identical blocks, each containing three layers: the syntactic attention layer, densely connected layer, and linear combination layer. The syntactic attention layer is the innovative part of SDANN.

As discussed in Section 1, typical pruning methods are preset and transform the full tree into a subtree, forming the basis of the adjacency matrix. This can be likened to hard attention (Xu *et al.* 2015), where non-subtree edges are neglected. Our model employs a 'soft pruning' method in the syntactic attention layer, assigning weights to all edges, learned end-to-end.

In this layer, the original tree is converted into a fully connected edge-weighted graph by constructing a syntactic attention adjacency matrix $\tilde{\mathbf{A}}$, where each entry $\tilde{\mathbf{A}}_{ij}$ represents the weight from node i to node j. Figure **??** illustrates this with $\tilde{\mathbf{A}}^{(\mathbf{1})}$ corresponding to graph $G^{(1)}$. We use a self-attention mechanism (Cheng *et al.* 2016), capturing interactions within a sequence, to construct $\tilde{\mathbf{A}}$. For computational layers, $\tilde{\mathbf{A}}$ replaces the original $\mathbf{A}$, without additional computational cost. This layer leverages attention to infer node relations, particularly for indirect paths, enabling the model to capture nuanced connections through differentiable functions.

Multi-head attention (Vaswani *et al.* 2017) is used for calculating $\tilde{\mathbf{A}}$, enabling simultaneous attention to different representation subspaces. The calculation involves a query and key-value pairs, with output as a weighted sum of values, based on the query and corresponding key.

$$\tilde{\mathbf{A}}^{(\mathbf{t})} = softmax(\frac{Q\mathbf{W}_i^Q \times (K\mathbf{W}_i^K)^T}{\sqrt{d}}) \tag{2}$$

Here, Q and K are equal to $\mathbf{h}^{(l-1)}$, the collective representation at layer $l-1$. The projections are parameter matrices $\mathbf{W}_i^Q, \mathbf{W}_i^K \in \mathbb{R}^{d\times d}$.

$\tilde{\mathbf{A}}^{(\mathbf{t})}$ is the t-th syntactic attention adjacency matrix corresponding to the t-th head, with up to N matrices constructed, where N is a hyper-parameter.

3.3. Linear Layer

In contrast to traditional pruning, resulting in smaller structures, our syntactic attention layer outputs a larger fully connected graph. Adapting from (Guo *et al.* 2019), we integrate dense connections (Huang *et al.* 2017) into SDANN to encapsulate more structural information on large graphs. These connections enable a deeper model, capturing rich local and non-local information for an improved graph representation.

Dense connectivity introduces direct connections from any layer to all preceding layers. We define $\mathbf{g}_j^{(l)}$ as the concatenation of the initial node representation and the representations from layers 1 to $l-1$:

$$\mathbf{g}_j^{(l)} = [\mathbf{x}_j; \mathbf{h}_j^{(1)}; ...; \mathbf{h}_j^{(l-1)}]. \tag{3}$$

Each densely connected layer has L sub-layers, with dimensions $d_{hidden} = d/L$. For instance, a layer with 3 sub-layers and 300 input dimensions will have $d_{hidden} = 100$, and the output dimension remains 300 (3×100). This structure, akin to DenseNets (Huang *et al.* 2017), improves parameter efficiency.

Given N different syntactic attention adjacency matrices, N separate densely connected layers are needed. The computation for each layer, corresponding to the t-th matrix $\tilde{\mathbf{A}}^{(\mathbf{t})}$, is modified as follows:

$$\mathbf{h}_{t_i}^{(l)} = \rho\Big(\sum_{j=1}^{n} \tilde{\mathbf{A}}_{ij}^{(t)} \mathbf{W}_t^{(l)} \mathbf{g}_j^{(l)} + \mathbf{b}_t^{(l)}\Big) \tag{4}$$

where $t = 1, ..., N$, selecting the appropriate weight matrix and bias term for each $\tilde{\mathbf{A}}^{(\mathbf{t})}$. The column dimension of $\mathbf{W}_t^{(l)}$ increases by d_{hidden} per sub-layer.

3.4. Linear Combination Layer

SDANN includes a linear combination layer to integrate representations from the N densely connected layers. The output is defined as:

$$\mathbf{h}_{comb} = \mathbf{W}_{comb}\mathbf{h}_{out} + \mathbf{b}_{comb} \tag{5}$$

where $\mathbf{h}_{out}$ is the concatenation of outputs from N layers, and $\mathbf{W}_{comb} \in \mathbb{R}^{(d\times N)\times d}$ and $\mathbf{b}_{comb}$ are the weight matrix and bias vector for the linear transformation.

3.5. SDANNs for Relation Extraction

Applying SDANN over the dependency tree yields hidden representations for all tokens. The goal of relation extraction is to predict relations among entities using these representations. Following (Zhang *et al.* 2018), we concatenate sentence and entity representations for the final classification. Sentence representation h_{sent} is obtained by:

$$h_{sent} = f(\mathbf{h}_{\mathbf{mask}}) = f(\text{SDANN}(\mathbf{x})) \tag{6}$$

where $\mathbf{h}_{\mathbf{mask}}$ represents selected non-entity token representations, and f is a max pooling function mapping n output vectors to one sentence vector. Entity representations are similarly obtained and concatenated with h_{sent}. A feed-forward neural network (FFNN) processes these representations for final prediction:

$$h_{final} = \text{FFNN}([h_{sent}; h_{e_1}; ...; h_{e_i}]) \tag{7}$$

h_{final} is then used in a logistic regression classifier for prediction.

4. Experiments

4.1. Experimental Framework

We assess the SDANN model's capabilities in two key areas: cross-sentence n-ary relation extraction and sentence-level relation extraction.

In cross-sentence n-ary relation extraction, we employ the dataset from (Peng *et al.* 2017), featuring 6,987 ternary and 6,087 binary relation instances from PubMed. Instances span multiple sentences and are categorized into five labels: resistance or nonresponse", sensitivity", response", resistance", and None". We bifurcate our analysis into binary-class n-ary relation extraction and multi-class n-ary relation extraction, with the former consolidating four relation types into Yes" and assigning None" as No", following (Peng *et al.* 2017).

For sentence-level relation extraction, the TACRED dataset (Zhang *et al.* 2017) and Semeval-10 Task 8 (Hendrickx *et al.* 2010) are utilized as per the methodologies in (Zhang *et al.* 2018). TACRED, with over 106K instances, enumerates 41 relation types plus a no relation" category, while Semeval-10 Task 8 offers 10,717 instances across 9 relations plus an other" category.

Hyper-parameter optimization is guided by development set outcomes. For cross-sentence tasks, we use the data split from (Song *et al.* 2018a)??, and for sentence-level tasks, we follow the development set guidelines from (Zhang *et al.* 2018). The number of attention heads N, block numbers M, and sub-layer counts L in densely connected layers are selected from predefined sets. Optimal combinations are identified as (N=2, M=2, L_1=2, L_2=4, d_{hidden}=340) for cross-sentence tasks and (N=3, M=2, L_1=2, L_2=4, d_{hidden}=300) for sentence-level tasks. GloVe vectors (Pennington *et al.* 2014) serve as initial word embeddings.

Model evaluations align with the metrics used in prior studies (Zhang *et al.* 2018; Song *et al.* 2018a). For cross-sentence tasks, test accuracy is reported as an average over five validation folds (Song *et al.*

2018a). In sentence-level tasks, we report micro-averaged F1 scores for TACRED and macro-averaged F1 scores for SemEval.

4.2. Results on Cross-Sentence *n*-ary Relation Extraction

In evaluating the SDANN model for cross-sentence *n*-ary relation extraction, we benchmark against three categories of models: 1) Feature-based classifiers (Quirk and Poon 2017) utilizing shortest dependency paths, 2) Graph-structured LSTM frameworks like Graph LSTM (Peng *et al.* 2017), Bidirectional DAG LSTM (Bidir DAG LSTM) (Song *et al.* 2018a), and Graph State LSTM (GS GLSTM) (Song *et al.* 2018a), which encode graphs from sentences with dependency edges, and 3) Graph Convolutional Networks (GCN) with pruned trees(Zhang *et al.* 2018). Additionally, the tree-structured LSTM method (SPTree)(Miwa and Bansal 2016) is included for binary relation extraction of drug-mutations, as per(Song *et al.* 2018a). Results are presented in Table 1.

Table 1. Efficacy of SDANN in binary-class and multi-class *n*-ary relation extraction: A comparative analysis. T" indicates ternary interactions and B" binary interactions. `Single` covers single-sentence instances, while `Cross` includes all instances. *K* in GCN models refers to the pruning range from the dependency path in the LCA subtree.

Model	**Binary-class**				**Multi-class**	
	T		B		T	B
	`Single`	`Cross`	`Single`	`Cross`	`Cross`	`Cross`
Feature-Based (Quirk and Poon 2017)	74.7	77.7	73.9	75.2	-	-
SPTree (Miwa and Bansal 2016)	-	-	75.9	75.9	-	-
Graph LSTM-EMBED (Peng *et al.* 2017)	76.5	80.6	74.3	76.5	-	-
Graph LSTM-FULL (Peng *et al.* 2017)	77.9	80.7	75.6	76.7	-	-
+ multi-task	-	82.0	-	78.5	-	-
Bidir DAG LSTM (Song *et al.* 2018a)	75.6	77.3	76.9	76.4	51.7	50.7
GS GLSTM (Song *et al.* 2018a)	80.3	83.2	83.5	83.6	71.7	71.7
GCN (Full Tree) (Zhang *et al.* 2018)	84.3	84.8	84.2	83.6	77.5	74.3
GCN (*K*=0) (Zhang *et al.* 2018)	85.8	85.8	82.8	82.7	75.6	72.3
GCN (*K*=1) (Zhang *et al.* 2018)	85.4	85.7	83.5	83.4	78.1	73.6
GCN (*K*=2) (Zhang *et al.* 2018)	84.7	85.0	83.8	83.7	77.9	73.1
SDANN (ours)	**87.1**	**87.0**	**85.2**	**85.6**	**79.7**	**77.4**

Focusing on binary-class *n*-ary relation extraction, the SDANN model demonstrates superior performance, achieving 87.1% and 87.0% accuracies for single sentence (`Single`) and all instances (`Cross`) settings in ternary relation extraction. This outperforms all baselines, including surpassing the GS GLSTM model by significant margins. In binary relation extraction, SDANN continues to excel, consistently outshining both GS GLSTM and GCN models.

The SDANN model's effectiveness is attributed to its advanced graph convolution capabilities, which surpass traditional full tree-based methods like GS GLSTM. This improvement is likely due to the synergy between the densely connected layers and attention guided layers in SDANN, which enhance long-distance dependency learning without pruning, and refine the information extraction process.

For the multi-class classification task, our SDANN model maintains its lead, outperforming GS GLSTM and all GCN models. This further underscores SDANN's proficiency in handling complex relation extraction tasks.

4.3. Results on Sentence-level Relation Extraction

Turning to the TACRED dataset for sentence-level relation extraction, we compare SDANN against both dependency-based and sequence-based models. Dependency-based contenders include logistic regression classifiers (LR) (Zhang *et al.* 2017), Shortest Path LSTM (SDP-LSTM)(Xu *et al.* 2015), Tree-LSTM (Tai *et al.* 2015), and GCN variants (Zhang *et al.* 2018). The sequence-based category is represented by Position Aware LSTM (PA-LSTM)(Zhang *et al.* 2017). The results are detailed in Table 2.

The SDANN model showcases its efficacy by surpassing the GCN in F1 score. The integration of contextualized information via a bi-directional LSTM network, resulting in the Contextualized SDANN (C-SDANN), further elevates its performance, outdoing the C-GCN model. This enhancement signifies the importance of contextual data in relation extraction tasks.

Table 2. Performance analysis on the TACRED dataset. Asterisked models' results are as reported in respective citations. The default seed (0) C-SDANN achieves 69.0 F1 score. Extended results over multiple runs are provided.

Model	P	R	F1
LR (Zhang *et al.* 2017)	**73.5**	49.9	59.4
SDP-LSTM (Xu *et al.* 2015)*	66.3	52.7	58.7
Tree-LSTM (Tai *et al.* 2015)**	66.0	59.2	62.4
PA-LSTM (Zhang *et al.* 2017)	65.7	**64.5**	65.1
GCN (Zhang *et al.* 2018)	69.8	59.0	64.0
C-GCN (Zhang *et al.* 2018)	69.9	63.3	66.4
SDANN (ours)	69.9	60.9	65.1
C-SDANN (ours)	73.1	64.2	**69.0**

4.4. Results on Sentence-level Relation Extraction

In the realm of sentence-level relation extraction, our SDANN model was scrutinized against two distinct categories of models on the TACRED dataset. The first category included dependency-based models such as the Logistic Regression Classifier (LR) (Zhang *et al.* 2017), Shortest Path LSTM (SDP-LSTM)(Xu *et al.* 2015), Tree-structured LSTM (Tree-LSTM)(Tai *et al.* 2015), and Graph Convolutional Networks (GCN) including Contextualized GCN (C-GCN) (Zhang *et al.* 2018). The second category was represented by sequence-based models, notably the Position Aware LSTM (PA-LSTM)(Zhang *et al.* 2017). The results of this comparative analysis are detailed in Table2.

The SDANN model, in its C-SDANN form (an extension with a bi-directional LSTM for contextual representation), demonstrated significant efficacy, achieving a noteworthy F1 score of 69.0, surpassing the state-of-the-art C-GCN model. This performance indicates the effectiveness of integrating contextual information in the SDANN framework for nuanced relation extraction tasks.

Further evaluation on the SemEval dataset, under the same experimental conditions as (Zhang *et al.* 2018), revealed consistent superiority of our C-SDANN model. Despite the smaller size of the SemEval dataset compared to TACRED, C-SDANN achieved an F1 score of 85.7, clearly outperforming the C-GCN model. These results, presented in Table 3, underscore the model's robust generalizability across different dataset scales and complexities.

Table 3. Performance of C-SDANN on the SemEval dataset, highlighting its superior generalizability.

Model	F1
SVM (Rink and Harabagiu 2010)	82.2
SDP-LSTM (Xu *et al.* 2015)	83.7
SPTree (Miwa and Bansal 2016)	84.4
PA-LSTM (Zhang *et al.* 2017)	82.7
C-GCN (Zhang *et al.* 2018)	84.8
C-SDANN (ours)	**85.7**

4.5. Further Results

Ablation Study.

In our ablation study of the SDANN model, specifically its contextualized variant, we explored the impact of removing key components like attention-guided and densely connected layers. The results, presented in Table 4, illustrate that both types of layers significantly contribute to the model's performance. The attention-guided layer, in particular, shows a more pronounced effect, indicating its crucial role in enhancing the model's capabilities.

Table 4. Ablation study for the contextualized SDANN model.

Model	F1
SDANN (contextualized)	69.0
– Attention-guided layer	67.1
– Densely connected layer	67.3
– Both attention and densely connected layers	66.7
– Feed-Forward layer	67.8

Performance with Pruned Trees.

The effectiveness of SDANN under varying levels of tree pruning is detailed in Table 5. Notably, SDANN achieves its best performance with the full tree structure, surpassing the versions with pruned trees. This underscores the advantage of the model's ability to utilize comprehensive syntactic information, as opposed to relying on pruning strategies.

Table 5. Performance of SDANN with various levels of tree pruning.

Model	F1
SDANN (Full tree structure)	69.0
SDANN (Pruning level K=2)	67.5
SDANN (Pruning level K=1)	67.9
SDANN (Pruning level K=0)	67.0

Performance against Sentence Length.

Our analysis across different sentence lengths demonstrated that the full tree SDANN model consistently outperforms its pruned counterparts and the C-GCN model. The superiority of the full tree structure becomes more evident as sentence length increases, highlighting the model's proficiency in handling more complex dependency structures.

Performance against Training Data Size.

Under various training data size scenarios, the SDANN model consistently surpassed the C-GCN model, even with lesser training data. This efficiency becomes more apparent with increasing amounts of data, showcasing SDANN's effectiveness in resource utilization and learning from extensive data sets.

5. Conclusion and Future Directions

In this paper, we present a groundbreaking model, Syntactic Dependency-Aware Neural Networks (SDANN), which has demonstrated remarkable success in diverse relation extraction tasks, surpassing previous state-of-the-art models. The SDANN framework is unique in its methodology; it processes entire syntactic dependency trees, effectively sifting through and harnessing valuable information in an end-to-end manner. This contrasts sharply with prior methods that often relied on partial tree structures. Looking ahead, there are several promising directions for further exploration with SDANN.

A particularly intriguing avenue is to explore how this innovative framework can be harnessed to enhance graph representation learning in various graph-related tasks, as discussed in (Bastings *et al.* 2017). The potential of SDANN in such applications is vast, given its ability to capture complex relational data from full syntactic structures, which opens up new possibilities in graph-based learning paradigms.

References

Quirk, C.; Poon, H. Distant Supervision for Relation Extraction beyond the Sentence Boundary. Proc. of EACL, 2017.

Zhang, Y.; Zhong, V.; Chen, D.; Angeli, G.; Manning, C.D. Position-aware Attention and Supervised Data Improve Slot Filling. Proc. of EMNLP, 2017.

Yu, M.; Yin, W.; Hasan, K.S.; dos Santos, C.N.; Xiang, B.; Zhou, B. Improved Neural Relation Detection for Knowledge Base Question Answering. Proc. of ACL, 2017.

Fei, H.; Li, J.; Wu, S.; Li, C.; Ji, D.; Li, F. Global Inference with Explicit Syntactic and Discourse Structures for Dialogue-Level Relation Extraction. Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI, 2022, pp. 4082–4088.

Wu, S.; Fei, H.; Ren, Y.; Ji, D.; Li, J. Learn from Syntax: Improving Pair-wise Aspect and Opinion Terms Extraction with Rich Syntactic Knowledge. Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, 2021, pp. 3957–3963.

Xiang, C.; Zhang, J.; Li, F.; Fei, H.; Ji, D. A semantic and syntactic enhanced neural model for financial sentiment analysis. *Information Processing & Management* **2022**, *59*, 102943.

Fei, H.; Zhang, Y.; Ren, Y.; Ji, D. Latent Emotion Memory for Multi-Label Emotion Classification. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, pp. 7692–7699.

Tang, D.; Qin, B.; Feng, X.; Liu, T. Effective LSTMs for Target-Dependent Sentiment Classification. Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, 2016, pp. 3298–3307.

Wu, S.; Fei, H.; Li, F.; Zhang, M.; Liu, Y.; Teng, C.; Ji, D. Mastering the Explicit Opinion-Role Interaction: Syntax-Aided Neural Transition System for Unified Opinion Role Labeling. Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence, 2022, pp. 11513–11521.

Ma, D.; Li, S.; Zhang, X.; Wang, H. Interactive Attention Networks for Aspect-Level Sentiment Classification. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, 2017, pp. 4068–4074.

Fei, H.; Zhang, M.; Ji, D. Cross-Lingual Semantic Role Labeling with High-Quality Translated Training Corpus. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 7014–7026.

Chen, P.; Sun, Z.; Bing, L.; Yang, W. Recurrent attention network on memory for aspect sentiment analysis. Proceedings of the 2017 conference on empirical methods in natural language processing, 2017, pp. 452–461.

Fei, H.; Zhang, M.; Li, B.; Ji, D. End-to-end Semantic Role Labeling with Neural Transition-based Model. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, pp. 12803–12811.

Zhang, M.; Qian, T. Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis. Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP), 2020, pp. 3540–3549.

Wu, S.; Fei, H.; Ren, Y.; Li, B.; Li, F.; Ji, D. High-Order Pair-Wise Aspect and Opinion Terms Extraction With Edge-Enhanced Syntactic Graph Convolution. *IEEE ACM Trans. Audio Speech Lang. Process.* **2021**, *29*, 2396–2406.

Chen, C.; Teng, Z.; Zhang, Y. Inducing target-specific latent structures for aspect sentiment classification. Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP), 2020, pp. 5596–5607.

Fei, H.; Wu, S.; Ren, Y.; Zhang, M. Matching Structure for Dual Learning. Proceedings of the International Conference on Machine Learning, ICML, 2022, pp. 6373–6391.

Huang, L.; Sun, X.; Li, S.; Zhang, L.; Wang, H. Syntax-aware graph attention network for aspect-level sentiment classification. Proceedings of the 28th international conference on computational linguistics, 2020, pp. 799–810.

Fei, H.; Li, F.; Li, B.; Ji, D. Encoder-Decoder Based Unified Semantic Role Labeling with Label-Aware Syntax. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, pp. 12794–12802.

Fei, H.; Ren, Y.; Zhang, Y.; Ji, D. Nonautoregressive Encoder-Decoder Neural Framework for End-to-End Aspect-Based Sentiment Triplet Extraction. *IEEE Transactions on Neural Networks and Learning Systems* **2023**, *34*, 5544–5556.

Hou, X.; Huang, J.; Wang, G.; He, X.; Zhou, B. Selective attention based graph convolutional networks for aspect-level sentiment classification. *arXiv preprint arXiv:1910.10857* **2019**.

Fei, H.; Li, J.; Ren, Y.; Zhang, M.; Ji, D. Making Decision like Human: Joint Aspect Category Sentiment Analysis and Rating Prediction with Fine-to-Coarse Reasoning. Proceedings of the ACM Web Conference 2022, WWW, 2022, pp. 3042–3051.

Zeng, D.; Liu, K.; Lai, S.; Zhou, G.; Zhao, J. Relation Classification via Convolutional Deep Neural Network. Proc. of COLING, 2014.

Wang, L.; Cao, Z.; de Melo, G.; Liu, Z. Relation Classification via Multi-Level Attention CNNs. Proc. of ACL, 2016.

Bunescu, R.C.; Mooney, R.J. A Shortest Path Dependency Kernel for Relation Extraction. Proc. of EMNLP, 2005.

Peng, N.; Poon, H.; Quirk, C.; Toutanova, K.; tau Yih, W. Cross-Sentence N-ary Relation Extraction with Graph LSTMs. *Transactions of the Association for Computational Linguistics* **2017**, *5*, 101–115.

Xu, K.; Feng, Y.; Huang, S.; Zhao, D. Semantic Relation Classification via Convolutional Neural Networks with Simple Negative Sampling. Proc. of EMNLP, 2015.

Fei, H.; Wu, S.; Li, J.; Li, B.; Li, F.; Qin, L.; Zhang, M.; Zhang, M.; Chua, T.S. LasUIE: Unifying Information Extraction with Latent Adaptive Structure-aware Generative Language Model. Proceedings of the Advances in Neural Information Processing Systems, NeurIPS 2022, 2022, pp. 15460–15475.

Xu, Y.; Mou, L.; Li, G.; Chen, Y.; Peng, H.; Jin, Z. Classifying Relations via Long Short Term Memory Networks along Shortest Dependency Paths. Proc. of EMNLP, 2015.

Miwa, M.; Bansal, M. End-to-End Relation Extraction using LSTMs on Sequences and Tree Structures. Proc. of ACL, 2016.

Zhang, Y.; Qi, P.; Manning, C.D. Graph Convolution over Pruned Dependency Trees Improves Relation Extraction. Proc. of EMNLP, 2018.

Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. Proc. of ICLR, 2017.

Li, J.; Fei, H.; Liu, J.; Wu, S.; Zhang, M.; Teng, C.; Ji, D.; Li, F. Unified Named Entity Recognition as Word-Word Relation Classification. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, pp. 10965–10973.

Fei, H.; Ren, Y.; Zhang, Y.; Ji, D.; Liang, X. Enriching contextualized language model from knowledge graph for biomedical information extraction. *Briefings in Bioinformatics* **2021**, *22*.

Dong, L.; Wei, F.; Tan, C.; Tang, D.; Zhou, M.; Xu, K. Adaptive recursive neural network for target-dependent twitter sentiment classification. Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 2: Short papers), 2014, Vol. 2, pp. 49–54.

Fei, H.; Ren, Y.; Ji, D. Boundaries and edges rethinking: An end-to-end neural model for overlapping entity relation extraction. *Information Processing & Management* **2020**, *57*, 102311.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. Proc. of NeurIPS, 2017.

Fei, H.; Ji, D.; Li, B.; Liu, Y.; Ren, Y.; Li, F. Rethinking Boundaries: End-To-End Recognition of Discontinuous Mentions with Pointer Networks. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, pp. 12785–12793.

Mukherjee, R.; Shetty, S.; Chattopadhyay, S.; Maji, S.; Datta, S.; Goyal, P. Reproducibility, replicability and beyond: Assessing production readiness of aspect based sentiment analysis in the wild. Advances in Information Retrieval: 43rd European Conference on IR Research, ECIR 2021, Virtual Event, March 28–April 1, 2021, Proceedings, Part II 43. Springer, 2021, pp. 92–106.

Fei, H.; Chua, T.; Li, C.; Ji, D.; Zhang, M.; Ren, Y. On the Robustness of Aspect-based Sentiment Analysis: Rethinking Model, Data, and Training. *ACM Transactions on Information Systems* **2023**, *41*, 50:1–50:32.

Zhuang, L.; Fei, H.; Hu, P. Knowledge-enhanced event relation extraction via event ontology prompt. *Inf. Fusion* **2023**, *100*, 101919.

Zhang, C.; Li, Q.; Song, D. Aspect-based sentiment classification with aspect-specific graph convolutional networks. *arXiv preprint arXiv:1909.03477* **2019**.

Fei, H.; Wu, S.; Ren, Y.; Li, F.; Ji, D. Better Combine Them Together! Integrating Syntactic Constituency and Dependency Representations for Semantic Role Labeling. Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, 2021, pp. 549–559.

Chen, X.; Sun, C.; Wang, J.; Li, S.; Si, L.; Zhang, M.; Zhou, G. Aspect sentiment classification with document-level sentiment preference modeling. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 3667–3677.

Li, J.; Xu, K.; Li, F.; Fei, H.; Ren, Y.; Ji, D. MRN: A Locally and Globally Mention-Based Reasoning Network for Document-Level Relation Extraction. Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, 2021, pp. 1359–1370.

Liu, J.; Fei, H.; Li, F.; Li, J.; Li, B.; Zhao, L.; Teng, C.; Ji, D. TKDP: Threefold Knowledge-enriched Deep Prompt Tuning for Few-shot Named Entity Recognition. *CoRR* **2023**, *abs/2306.03974*.

Pontiki, M.; Galanis, D.; Papageorgiou, H.; Androutsopoulos, I.; Manandhar, S.; AL-Smadi, M.; Al-Ayyoub, M.; Zhao, Y.; Qin, B.; De Clercq, O.; others. Semeval-2016 task 5: Aspect based sentiment analysis. ProWorkshop on Semantic Evaluation (SemEval-2016). Association for Computational Linguistics, 2016, pp. 19–30.

Fei, H.; Li, F.; Li, C.; Wu, S.; Li, J.; Ji, D. Inheriting the Wisdom of Predecessors: A Multiplex Cascade Framework for Unified Aspect-based Sentiment Analysis. Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI, 2022, pp. 4096–4103.

Wang, F.; Li, F.; Fei, H.; Li, J.; Wu, S.; Su, F.; Shi, W.; Ji, D.; Cai, B. Entity-centered Cross-document Relation Extraction. Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 2022, pp. 9871–9881.

Fei, H.; Ren, Y.; Ji, D. Retrofitting Structure-aware Transformer Language Model for End Tasks. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, 2020, pp. 2151–2161.

Jia, Y.; Wang, Y.; Zan, H.; Xie, Q. Syntactic information and multiple semantic segments for aspect-based sentiment classification. *International Journal of Asian Language Processing* **2021**, *31*, 2250006.

Cao, H.; Li, J.; Su, F.; Li, F.; Fei, H.; Wu, S.; Li, B.; Zhao, L.; Ji, D. OneEE: A One-Stage Framework for Fast Overlapping and Nested Event Extraction. Proceedings of the 29th International Conference on Computational Linguistics, 2022, pp. 1953–1964.

Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. Proc. of CVPR, 2017.

Guo, Z.; Zhang, Y.; Teng, Z.; Lu, W. Densely Connected Graph Convolutional Networks for Graph-to-Sequence Learning. *Transactions of the Association of Computational Linguistics* **2019**.

Tai, K.S.; Socher, R.; Manning, C.D. Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks. Proc. of ACL, 2015.

Zelenko, D.; Aone, C.; Richardella, A. Kernel Methods for Relation Extraction. Proc. of EMNLP, 2002.

McDonald, R.T.; Pereira, F.; Kulick, S.; Winters, R.S.; Jin, Y.; White, P.S. Simple Algorithms for Complex Relation Extraction with Applications to Biomedical IE. Proc. of ACL, 2005.

Mintz, M.; Bills, S.; Snow, R.; Jurafsky, D. Distant supervision for relation extraction without labeled data. Proc. of ACL, 2009.

Nguyen, T.H.; Grishman, R. Relation Extraction: Perspective from Convolutional Neural Networks. Proc. of VS@NAACL-HLT, 2015.

Zhou, P.; Shi, W.; Tian, J.; Qi, Z.; Li, B.; Hao, H.; Xu, B. Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification. Proc. of ACL, 2016.

Fei, H.; Liu, Q.; Zhang, M.; Zhang, M.; Chua, T.S. Scene Graph as Pivoting: Inference-time Image-free Unsupervised Multimodal Machine Translation with Visual Scene Hallucination. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023, pp. 5980–5994.

doi:10.20944/preprints202311.1681.v1

Vu, N.T.; Adel, H.; Gupta, P.; Schütze, H. Combining Recurrent and Convolutional Neural Networks for Relation Classification. Proc. of NAACL-HLT, 2016.

Verga, P.; Strubell, E.; McCallum, A. Simultaneously Self-Attending to All Mentions for Full-Abstract Biological Relation Extraction. Proc. of NAACL-HLT, 2018.

Song, L.; Zhang, Y.; Wang, Z.; Gildea, D. N-ary Relation Extraction using Graph State LSTM. Proc. of EMNLP, 2018.

Song, L.; Zhang, Y.; Wang, Z.; Gildea, D. A Graph-to-Sequence Model for AMR-to-Text Generation. Proc. of ACL, 2018.

Fei, H.; Zhang, M.; Zhang, M.; Chua, T.S. Constructing Code-mixed Universal Dependency Forest for Unbiased Cross-lingual Relation Extraction. Findings of the Association for Computational Linguistics: ACL 2023, 2023, pp. 9395–9408.

Fei, H.; Ren, Y.; Ji, D. Mimic and Conquer: Heterogeneous Tree Structure Distillation for Syntactic NLP. Findings of the Association for Computational Linguistics: EMNLP 2020, 2020, pp. 183–193.

Liu, Y.; Wei, F.; Li, S.; Ji, H.; Zhou, M.; Wang, H. A Dependency-Based Neural Network for Relation Classification. Proc. of ACL, 2015.

Gori, M.; Monfardini, G.; Scarselli, F. A new model for learning in graph domains. Proc. of IJCNN, 2005.

Bruna, J. Spectral Networks and Deep Locally Connected Networks on Graphs. Proc. of ICLR, 2014.

Henaff, M.; Bruna, J.; LeCun, Y. Deep Convolutional Networks on Graph-Structured Data. *arXiv preprint* **2015**.

Defferrard, M.; Bresson, X.; Vandergheynst, P. Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering. Proc. of NeurIPS, 2016.

Fei, H.; Ren, Y.; Ji, D. Improving Text Understanding via Deep Syntax-Semantics Communication. Findings of the Association for Computational Linguistics: EMNLP 2020, 2020, pp. 84–93.

Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. Proc. of ICLR, 2018.

Marcheggiani, D.; Titov, I. Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling. Proc. of EMNLP, 2017.

Xu, K.; Ba, J.; Kiros, R.; Cho, K.; Courville, A.C.; Salakhutdinov, R.; Zemel, R.S.; Bengio, Y. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. Proc. of ICML, 2015.

Cheng, J.; Dong, L.; Lapata, M. Long Short-Term Memory-Networks for Machine Reading. Proc. of EMNLP, 2016.

Hendrickx, I.; Kim, S.N.; Kozareva, Z.; Nakov, P.; Séaghdha, D.Ó.; Padó, S.; Pennacchiotti, M.; Romano, L.; Szpakowicz, S. SemEval-2010 Task 8: Multi-Way Classification of Semantic Relations between Pairs of Nominals. SemEval@ACL, 2010.

Pennington, J.; Socher, R.; Manning, C.D. Glove: Global Vectors for Word Representation. Proc. of EMNLP, 2014.

Rink, B.; Harabagiu, S.M. UTD: Classifying Semantic Relations by Combining Lexical and Semantic Resources. SemEval@ACL, 2010.

Bastings, J.; Titov, I.; Aziz, W.; Marcheggiani, D.; Sima'an, K. Graph Convolutional Encoders for Syntax-aware Neural Machine Translation. EMNLP, 2017.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.