Journal of Data Intelligence, Vol. 3, No. 1 (2021) 101–115

SIGMALAW PBSA - A DEEP LEARNING APPROACH FOR ASPECT-BASED SENTIMENT ANALYSIS IN LEGAL OPINION TEXTS

ISANKA RAJAPAKSHA
Department of Computer Science & Engineering, University of Moratuwa
Sri Lanka
israjapaksha.16@cse.mrt.ac.lk

CHANIKA RUCHINI MUDALIGE
Department of Computer Science & Engineering, University of Moratuwa
Sri Lanka
chanikaruchini.16@cse.mrt.ac.lk

DILINI KARUNARATHNA
Department of Computer Science & Engineering, University of Moratuwa
Sri Lanka
dilinirasanjana.16@cse.mrt.ac.lk

NISANSA DE SILVA
Department of Computer Science & Engineering, University of Moratuwa
Sri Lanka
NisansaDdS@cse.mrt.ac.lk

GATHIKA RATNAYAKA
Department of Computer Science & Engineering, University of Moratuwa
Sri Lanka
gathika.14@cse.mrt.ac.lk

AMAL SHEHAN PERERA
Department of Computer Science & Engineering, University of Moratuwa
Sri Lanka
shehan@cse.mrt.ac.lk

When lawyers and legal officers are working on a new legal case, they are supposed have properly studied prior cases similar to the current case, as the prior cases can provide valuable information which can have a direct impact on the outcomes of the current court case. Therefore, developing methodologies which are capable of automatically extracting information from legal opinion texts related to previous court cases can be considered as an important tool when it comes to the legal technology ecosystem. In this study, we focus on finding advantageous and disadvantageous facts or arguments in court cases, which is one of the most critical and time-consuming tasks in court case analysis. The Aspect-based Sentiment Analysis concept is used as the base of this study to perform legal information extraction. In this paper, we introduce a solution to predict sentiment value of sentences in legal documents in relation to its legal parties. The proposed approach employs a fine-grained sentiment analysis (Aspect-Based Sentiment Analysis) technique to achieve this task. Sigmalaw PBSA is a novel deep learning-based model for ABSA which is specifically designed for legal opinion texts. We evaluate the Sigmalaw PBSA model and existing ABSA models on the SigmaLaw-ABSA dataset which consists of 2000 legal opinion texts fetched from a public online data base. Experiments show that our model outperforms the state-of-the-art models. We also conduct an ablation study to identify which methods are most effective for legal texts.

Keywords: Legal Information Extraction, Legal domain, Aspect-Based Sentiment Analysis, Deep learning, NLP

1. Introduction

Factual scenario analysis of previous court cases holds a significant importance to lawyers and legal officers whenever they are handling a new legal court case. Legal officials are expected to analyse previous court cases and statutes to find supporting arguments before they represent a client at a trial. As the number of legal cases increases, legal professionals typically endure heavy workloads on a daily basis, and they may become overwhelmed and as a result of that, be unable to obtain quality analysis. In this analysis process, identifying advantageous and disadvantageous statements relevant to legal parties [1–5] can be considered a critical and time consuming task. By automating this task, legal officers will be able to reduce their workload significantly. In this paper, we introduce a solution to predict sentiment value of sentences in legal documents in relation to its legal parties. The proposed approach employs a fine-grained sentiment analysis technique to achieve this task.

Opinion mining or sentiment analysis is identifying opinions and then classifying them into several polarity levels (Positive,Neutral or Negative) using computational linguistics and information retrieval [6]. When it comes to the legal domain, sentiment analysis becomes a much challenging area because of the domain-specific meanings or behavior of words in the legal opinion texts. The complexity and the length of the sentences also increases the difficulty. Sentiment Analysis (SA) can be divided into 4 levels; document-level sentiment analysis, sentence-level sentiment analysis, phrase-level sentiment analysis, and aspect level sentiment analysis [7]. Document Level SA considers the whole document is about an entity and classifies whether the expressed sentiment is positive, negative, or neutral; Sentence Level SA determines the sentiment of each sentence, Phrase Level SA [8] focus on finding out the sentiment of phrases; Entity or Aspect-Based SA performs finer-grained analysis in which all entities and their aspects should be extracted and the sentiment on them should also be determined [9].

Sentences in a legal case usually contain two or more members/entities which belong to main legal parties (*plaintiff*, *petitioner*, *defendant*, and *respondent*). Extracting opinions with respect to each legal party cannot be performed only by using document-level, sentence-level, or phrase-level sentiment analysis. Aspect-based sentiment analysis (ABSA) is the most appropriate and fine-grained solution to perform Party-Based Sentiment Analysis (PBSA) in the legal domain [1]. In aspect-based sentiment analysis, we can identify there processing steps such as ”identification, classification, and aggregation" [9]. Generally, in ABSA aspects are extracted from a given text and then each aspect is allocated a sentiment level (*positive*, *negative*, or *neutral*) [10]. The members of legal parties in a court case are considered as aspects and therefore performing ABSA in the legal opinion texts can also be termed as Party-Based Sentiment Analysis (PBSA) [1].

A number of studies have addressed Aspect-based Sentiment Analysis in different domains such as restaurants, hotels, movies, products reviews, government services, mobile phones and telecommunication [11]. When it comes to the legal domain, sentiment analysis becomes a challenging area because of the domain-specific meanings and behaviour of words in the legal opinion texts [12]. Languages being used are sometimes mixed (i.e., English, Latin, etc.) and in some situations, the meaning of the words and context varies from that of domain interpretations. The complexity structure and the length of the sentences also increase the difficulty. As a result of the above factors, it is difficult to obtain a comparative accuracy to other areas such as customer feedback, movie or product reviews, and political comments.

Example 1

- Sentence 1.1: *After obtaining a warrant, the officials searched Lee's house, where they found drugs, cash, and a loaded rifle.*

Example 1 contains a sentence extracted from Lee v. United States [13] which mentions two legal party members: *Lee* and *officials*. As the illegal materials were found at *Lee's house*, this sentence clearly shows a negative sentiment towards *Lee* and Positive sentiment towards *officials*.

Example 2

- Sentence 2.1: *Lee has demonstrated that he was prejudiced by his counsel's erroneous advice.*

Example 2 sentence is taken from Lee v. United States [13] and it consists of two party members; *Lee* and *counsel*.The sentence uses pronouns (*he*, *his*) referring to the person, *Lee*. According to the context of the sentence, both Lee and counsel have a negative sentiment. when determining the sentiment level of Lee, we need to take pronouns into account as pronouns of Lee also have influenced the sentiment values significantly.

The rule-based approach proposed by Rajapaksha et al. [1] can be identified as the first and only attempt to perform ABSA in the legal domain to the best of our knowledge. However, that approach has two weaknesses: (1) it significantly depends on the phrase-level sentiment annotator, (2) manually created rules may not cover all the sentence patterns. There are many existing deep-learning models with different architectures trained for different domains to fulfil a wide array of tasks. Despite that, to the best of our knowledge, there is no existing deep learning-based approach for ABSA in the legal domain. In this paper we show that, as the sentences in legal documents are often long and have a complex semantic structure, the existing model architectures, created for short sentences in general use, do not perform well for the legal domain. The main objective of this study is to propose a novel deep learning-based model (SigmaLaw-PBSA) for ABSA, designed specifically for the legal domain.

2. Related Work

Legal Domain can be considered as a domain which can be made more productive and effective with the introduction of Artificial Intelligence. This fact is evidenced by the number of research that have been conducted over many years on automatic legal information extraction. Past literature related to the legal domain covers the areas of information organization [14–16], information extraction [12, 17], and information retrieval [18]. Few studies have been published on tasks of legal jargon embeddings in vector spaces [12, 19].

2.1. *Legal Information Extraction*

When referring to the past literature, it is evident that within the legal domain, there exist very few studies related to sentiment analysis. The study by Gamage et al. [8] introduced a sentence-level sentiment annotator using transfer learning for the legal domain. In this proposed approach, the sentiment of a given sentence is classified into one of the two classes; *negative* and *non-negative*. But it does not take into consideration any party mentioned in the sentence when detecting the sentiment of the sentence. Moreover, the study by Ratnayaka et al. [20] have proposed methodologies to identify relationships among sentences in the legal documents. They have demonstrated that sentiment analysis can be used to identify sentences that provide different opinions on the same topic (contradictory opinions) within a legal opinion text. The study of Rajapaksha et al. [1] developed a rule-based approach which is built around a phrase level sentiment annotator [8] and manually created rules for sentiment detection of legal sentences with respect to legal parties. This can be identified as the first attempt to use ABSA in the legal domain.

2.2. *Existing Aspect-Based Sentiment Analysis Models*

From the early 2000s, Sentiment analysis has attracted a great deal of attention in natural language processing. That is because of the explosive growth of social media, which produces large amounts of opinionated data. Consequently, now it is an active research area in data mining, Web mining, and information retrieval. Since the term, *opinion*, is critical to several activities, the interest in sentiment analysis extends to many domains. Previous works on aspect-based sentiment analysis focus mainly on training sentiment classifiers based on bag-of-words feature extraction and features on sentiment lexicons [21]. Conventional representation methods consist of statistic based methods [22] and rule-based [23] methods. In past literature there can be seen statistical-based approaches which use MaxEnt-LDA [24] and SVM [22]. These techniques mainly depend on labor-intensive feature engineering and excess linguistic resources.

Lexicon-based approaches, machine learning-based approaches, and hybrids of machine learning and lexicon-based approaches are the main types of methods to perform Aspect-Based Sentiment Analysis (ABSA) [25]. Recently, deep neural network approaches have shown better results on aspect-based sentiment classification tasks due to its ability to generate the dense vectors of sentences without handcrafted features. RNN based approaches have shown better results on aspect-based sentiment classification tasks due to its ability to capture the sequential nature of languages. Further, RNN shows promising results compared to CNN in computing sequential data due to the ability to have memory on previous computations. However, a major drawback of simple RNN is having a vanishing gradient problem [26]. Hence, to overcome this limitation, Long-short term memory (LSTM) networks and gated recurrent units (GRU) have been introduced.

Tang et al. [27] proposed TD-LSTM which uses two Long Short-term Memory (LSTM) networks in order to extract important information from the left and right sides of the target. Although it improves the LSTM architecture, it is often impossible to distinguish between various sentiment polarities at a fine-grained level. A number of subsequent studies employed attention mechanisms to learn the key parts of sentences that should be given special focus in order to enhance the sentence representation. In that perspective, Wang et al. [28] proposed AT-LSTM and ATAE-LSTM, incorporating attention mechanisms to model relationships between aspects and context. In order to better understand target information, Cheng et al. [29] introduced the HiErarchical ATtention (HEAT) network with sentiment attention and aspect attention. Chen et al. [30] designed the RAM model by adopting multiple attentions to extract important information from memory. IAN, which was proposed by Ma et al. [31], utilizes a bidirectional attention mechanism and learns the attention for the contexts and the targets separately via interactive learning.

Although attention-based models have shown promising results over many ABSA tasks, they are not adequate to catch syntactic dependencies between aspect and the context words within the sentence. The important feature of Graph Convolutional Network (GCN) is that, it has the ability to draw syntactically related terms to the target aspect and then manipulate, multi-word associations and syntactical knowledge in long-range, utilizing GCN layers [32]. Zhao et al. [33] proposed ASGCN, adopting GCN for ABSA. Zhao et al. concluded that GCN improves overall efficiency by exploiting both syntactic knowledge and long-range word dependency. Zhao et al. [33] introduced the SDGCN model with the aim of modeling sentiment dependencies within a sentence among different target aspects.

3. Methodology

The ultimate goal of our proposed approach is to detect the sentiment polarity of sentences in legal texts with respect to each legal party mentioned in the sentence. While it is common for these sentiments to be diametrically opposed, it is not a universal rule. Legal texts usually consist of multiple legal parties having different inter-dependencies among them. Hence, the sentiment classifier should be developed to the level at which it is capable of classifying sentiment polarity values of multiple legal parties. In our approach, *positive*, *negative* and *neutral* are considered as sentiment polarities. The overall architecture of our proposed model is illustrated in Fig. 1. To perform the aspect sentiment classification, our model architecture is designed with the following layers; word embedding layer, Recurrent Neural Network (RNN) layer, position aware attention mechanism, GCN layer, and sentiment classification layer.

3.1. *Word Embedding Layer*

Word embedding layer maps each word to a high dimensional vector space. It is widely known that a strong word embedding is extremely important for composing a strong and efficient text representation for use at later stages. We used a pre-trained BERT (Bidirectional Encoder Representations from Transformers) model[a] [12] post-trained using the criminal court case legal opinion texts available in the SigmaLaw dataset to obtain the word embedding.

An input sentence (S), of N number of words is represented as $S = \{w_{s_1}, w_{s_2}, ..w_{s_N}\}$. A given sentence S, would include a set of aspect terms (S_a) of cardinality K where the ith aspect term is represented by A_i, $S_a = \{A_1, A_2, ..A_K\}$. Further, the ith aspect term, A_i, contains M_i number of words such that $M_i \in [1; N)$ represented by $A_i = \{w_{A_{i1}}, w_{A_{i2}}, ..w_{A_{iM_i}}\}$. By the virtue of aspects not overlapping

[a] Legal-BERT model - https://osf.io/s8dj6/

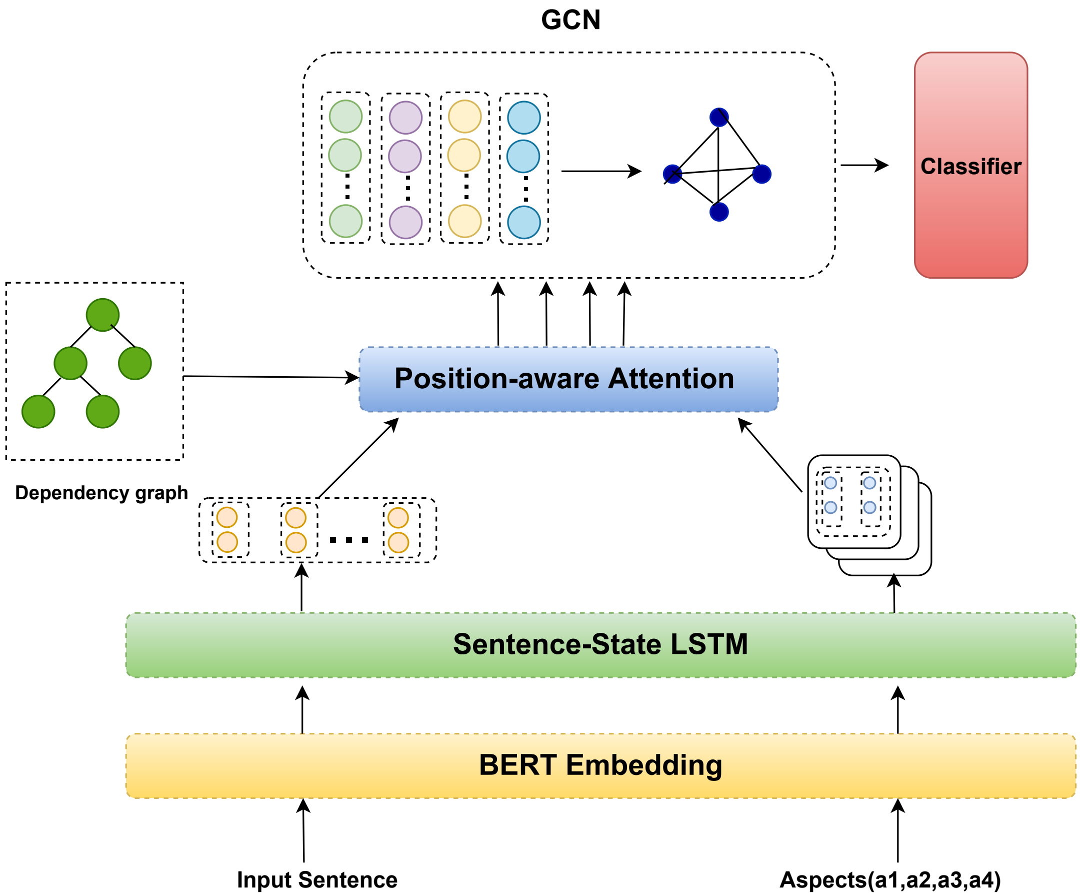

Fig. 1. Overall architecture

each other, $\sum_{i=1}^{K} M_i \leq N$ holds.

We use the above BERT model to get word embedding of the input sentence and all the aspect terms in the sentence. First, we construct the input as "[CLS] + input + [SEP]" and feed it to the BERT tokenizer. The special token [CLS] is added at the beginning of our text and the special token [SEP] is added to mark the end of a sentence. The BERT tokenizer then outputs tokens which correspond to BERT vocabulary. After mapping the token strings to their vocabulary indices, indexed tokens are next fed into the BERT model. Each word of the context and aspects are represented by a 768 dimensional embedding vector. The BERT model is used only for the word embedding purpose.

3.2. *RNN Layer*

In order to capture the contextual details for every word, on the top of the embedding layer we use Sentence-State LSTM (S-LSTM) [34]. Most of the existing model architectures use LSTM, Bi-LSTM, and Bi-GRU as the encoder. LSTM processes sequential data while maintaining long-term dependencies. However, when encoding long sentences the performance degrades. In our domain (legal documents), the sentences are comparatively longer than that of other domains. Therefore, aiming to address these limitations of existing deep-learning approaches, we leverage a sentence state LSTM (S-LSTM) to capture contextual information due to its proven performance [35]. Instead of sequentially processing words, the S-LSTM simultaneously models the hidden states of all words in each recurrent time stage.

After feeding the word embeddings of a sentence to the S-LSTM model, it returns the contextual state H^t of the sentence which consists of a sub hidden state h_i^t for each word w_i and a sentence-level sub hidden state s^t as shown in the equation 1.

$$H^t =< h_0^t, h_1^t, h_1^t, h_2^t, ..., h_{n-1}^t, s^t > \quad (1)$$

In our architecture, we use S-LSTM in order to get contextual hidden output of the sentence and contextual hidden outputs of aspects.

3.3. *Position Aware Attention Mechanism*

In a sentence, the sentiment polarity is heavily associated with the aspect-words and opinion terms of the sentence. Hence, the method that we adopt to rely on these aspect-terms is quite important in the process of sentiment analysis. The main weakness of RNN models is the inability to understand the most critical parts of the sentence for sentiment analysis. As a solution to this, we employ an attention mechanism which can grab the most important parts in a sentence. However, every word in a sentence is not equally important for determining sentiment polarity. Words which are closer to the target or having modifier relation to the target word should be given higher weights [36]. To ease this problem, we used an attention mechanism incorporating position information of each word in the sentence based on the current aspect term. We use position information here to incorporate the claim by He et al. [36] that the aspect sentiment polarity is mainly influenced by the context words that are situated very close to the target aspect.

Lee	was	found	guilty	because	the	attorney	had	provided	constitutionally	ineffective	assistance
[0,6]	[1,5]	[2,4]	[3,3]	[4,2]	[5,1]	[6,0]	[7,1]	[8,2]	[9,3]	[10,4]	[11,5]

Fig. 2. Basic relative distances to the aspects *Lee* and *attorney*

Here we used the bidirectional attention mechanism introduced by Zhao et al. [33] with two attention modules as context-to-aspect attention module and aspect-to-context attention module. We followed the same methodology for the calculation of attention weights. However, for position-aware representation, we used the distances along the dependency tree instead of the basic relative distances used in their approach. In our approach, as the distance, the length of the path from the specific word to the aspect in the dependency tree is used to encode the syntactic structure of the legal text. Fig. 2 illustrates the example sentence with basic relative distances to aspects and Fig. 3 shows distances along the dependency tree. When considering the two types of distances, we can see that the vital opinion words such as *guilty* and *ineffective* are closer to the relevant aspects in the Fig. 3 than in Fig. 2. The sentences in the court cases are comparatively much longer than other domains. Hence, opinion words are sometimes not close to the target. Therefore it is not suitable to get the basic relative distance between each word and the current aspect for position representation. The final output

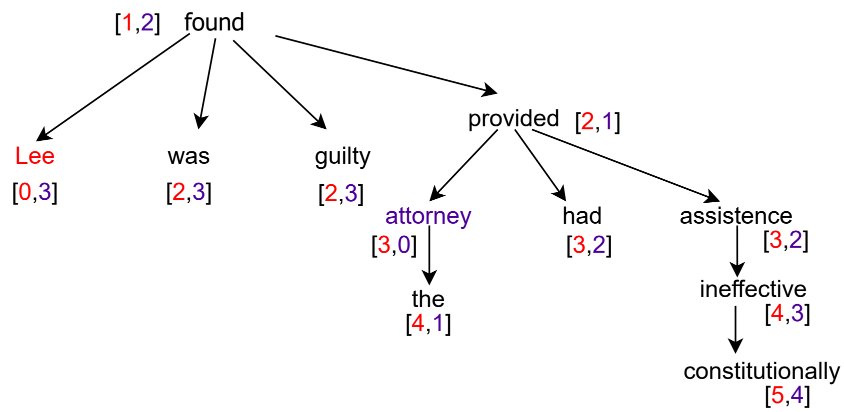

Fig. 3. Distances along the dependency tree to the aspects *Lee* and *attorney*

of the attention mechanism is the Aspect-specific representation between the target aspect (party) and context words given as $X = [x_1, x_2, .., x_K]$ where K denotes the number of aspects.

3.4. *Graph Convolution Network*

In order to capture the inter-dependencies between multiple aspects/parties in a sentence, we used GCN in our study following the observations reported by Zhao et al. [33] in their study. GCNs can be identified as a basic and efficient convolution neural network running on graphs which has the ability to collect interdependent knowledge from rich relational data. As the first stage of implementing the GCN layer, it is needed to construct a graph which we name as *Sentiment Graph*, where a node is a party (aspect) mentioned in the sentence and an edge is the inter-dependency relation between two nodes. If there is a dependency relationship between two parties in the sentence, we denote that by marking an edge between the corresponding two nodes. As shown in the Fig. 4, when creating the *Sentiment Graph*, we initially defined a fully-connected graph assuming that each aspect has a relationship with every other aspect of the sentence.

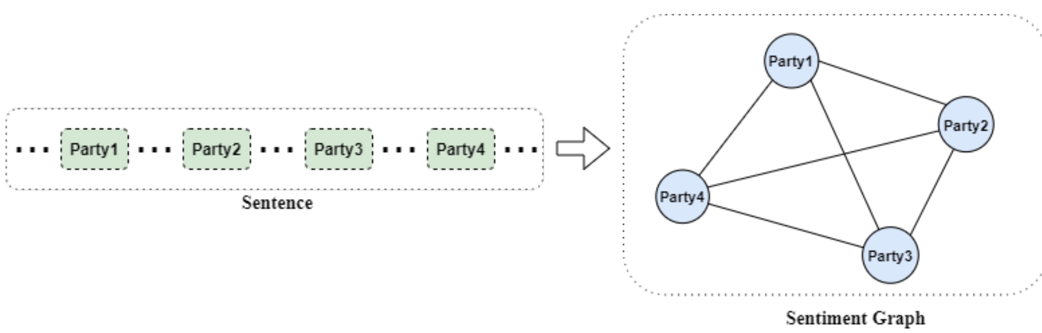

Fig. 4. Sentiment Graph

GCN generates a new vector representation for each node by discovering all relevant information about the neighboring nodes of the selected node. Moreover, when generating the new vector representation, it is needed to put attention on the information of the node itself. For that, we assume that each node has a self-loop. The new representation for a node can be defined as shown in the Eq. 2, where given v node, $N(v)$ defines the set of all neighbors of v, $W_{cross} \in \mathbb{R}^{d_m \times d_n}$, $W_{self} \in \mathbb{R}^{d_m \times d_n}$, $b_{cross} \in \mathbb{R}^{d_m \times 1}$, $b_{self} \in \mathbb{R}^{d_m \times 1}$, x_u is the uth aspect-specific representation taken from the output of attention layer.

$$x_v^1 = ReLU\left(\sum_{u \in N(v)} W_{cross} x_u + b_{cross}\right) + ReLU\left(W_{self} x_v + b_{self}\right) \tag{2}$$

We can expand the neighborhood for each node by stacking multiple GCN layers. As the input, each GCN layer gets the output form the previous layer and returns the new node representation. From the experiments, we identified that using more than two GCN layers reduces the accuracy. Therefore in our case, we use two GCN layers (see Eq. 3).

$$x_v^2 = ReLU\left(\sum_{u \in N(v)} W_{cross}^1 x_u^1 + b_{cross}^1\right) + ReLU\left(W_{self}^1 x_v^1 + b_{self}^1\right) \tag{3}$$

3.5. *Sentiment Classification*

Once the output of the GCN layer(x) is obtained, it is fed to a $Softmax$ layer to obtain a probability distribution over polarity decision space of C classes (where W and B are the learned weights and bias):

$$z = Softmax(Wx + b) \tag{4}$$

3.6. *Model Training*

The model is trained by the gradient descent algorithm with cross entropy loss and L2 regularization.

$$Loss = -\sum_{c=1}^{C} y \log \hat{y} + \lambda ||\theta||^2 \tag{5}$$

C denotes the number of classes (3 in our case), y is the true label, $\hat{y}$ is the predicted label, θ denotes all the parameters that need to regularized, and λ is the coefficient of L2-regularization.

4. Experiments

In this section we present and discuss the results that were obtained in relation to this work. We show that we obtain state-of-the-art results with comfortable margins.

4.1. *Data Set*

Experiments and evaluations were carried out on the SigmaLaw-ABSA [2] data set which consists of 2000 human-annotated legal sentences taken from previous court cases. The said court cases were originally fetched from the *SigmaLaw - Large Legal Text Corpus and Word Embedding data set* [12]. To the best of our knowledge SigmaLaw-ABSA is the only existing dataset for the Aspect-Based Sentiment Analysis in the legal domain. The dataset has been annotated by legal experts and it contains entities of different parties, their polarities, aspect category (*Petitioner* or *defendant*), and the category polarities. The data set has been designed to perform various research tasks in the legal domain

Table 1. Word embedding models comparison

Model	Accuracy	F1 score
GloVe [37]	0.6615	0.5798
BERT (base) [38]	0.6997	0.6193
BERT (legal domain) [39]	**0.7068**	**0.6281**

Table 2. Different word embedding strategies comparison of BERT model

Strategy	Accuracy	F1 score
Initial embedding	0.6670	0.5705
Last hidden layer	0.6921	0.6193
Concat last 4 layers	0.6987	0.6105
Sum all layers	0.6954	0.6098
Sum last 4 layers	**0.7086**	**0.6281**

including aspect extraction, polarity detection, aspect category identification, aspect category polarity detection. It is the only existing dataset for the aspect based sentiment analysis in the legal domain.

Legal sentence, members of legal parties in sentence, their polarities are the fields used for this study from the SigmaLaw-ABSA dataset. We feed the legal sentence as the input sentence and the legal party members as aspects into the BERT model. Polarity of the legal party members are used to evaluate the model.

4.2. *Parameter Setting*

For experiments, word embeddings for both context and targets are initialized by using 300-dimensional pretrained Glove word vectors and 760-dimensional Bert embeddings. Dimension of hidden state vectors of RNN is set to 300 and weights of the model are randomly initialized with uniform distribution. 600 is set as the output dimension of the GCN layer. We used Spacy[b] to calculate the distance through the dependency tree for attention mechanism and hidden states of the attention layer are set to 300. During the training, we set the batch size to 16, dropout to 0.1, coefficient of L2 is 10^{-5}, and used Adam optimizer with a learning rate of 0.001.

4.3. *Word Embedding Models Comparison*

In our experiments, we tried two word embedding methods: 300-dimensional GloVe [37] and BERT [38]. In BERT, two different BERT models were tried for the embedding layer: the base uncased English model and the pre-trained BERT model specially fine-tuned for the legal corpus. Table 1 shows the comparison of the results of above models. The legal-BERT model outperformed the other models. The BERT models use 12 layers of transformer encoders, and each output per token from each layer of these and initial input embedding can be used as a word embedding. We tried various vector combinations of hidden layers to get state-of-art results. Table 2 illustrates the result of various word-embedding strategies using the BERT model for legal domain.

4.4. *RNN Models Comparison*

[b] Spacy Toolkit - https://spacy.io/

Table 3. RNN models comparison

Model	Accuracy	F1 score
LSTM	0.6721	0.5964
Bi-LSTM	0.6987	0.6204
Bi-GRU	0.6854	0.6045
S-LSTM	**0.7086**	**0.6281**

Table 4. Performances of Different Models on SigmaLaw-ABSA

Model	Accuracy	F1 score
TD-LSTM [27]	0.6512	0.5647
TC-LSTM [27]	0.6182	0.5438
AE-LSTM [28]	0.6228	0.5588
AT-LSTM [28]	0.6272	0.5592
ATAE-LSTM [28]	0.6542	0.5802
IAN [31]	0.6332	0.5650
PBAN [40]	0.6332	0.5650
Cabasc [41]	0.6123	0.5643
RAM [30]	0.6639	0.6022
MemNet [42]	0.5389	0.4361
SDGCN [33]	0.6781	0.6121
ASDGCN [32]	0.6699	0.6001
SigmaLaw-PBSA	**0.7086**	**0.6281**

LSTM, Bi-LSTM, Bi-GRU and Sentence-State LSTM (S-LSTM) models were tested as the encoder for our approach as shown in Table 3. As S-LSTM offers richer contextual information exchange with more parallelism compared to BiLSTMs, it outperformed the other models. This is because it has strong representation power compared to the other RNNs [34]. This feature became relevant given that the sentences of Legal documents are often long and have a complex semantic structure.

4.5. *Overall Performance*

The experimental results generated on different existing models using the SigmaLaw-ABSA dataset [2] are shown in Table 4. By analysing the obtained results, we can conclude that, in the legal domain our proposed model outperforms every other existing model. We claim that it is mainly due to the complexity and the length of the sentences in the legal domain as it makes it difficult to those models to understand the sentence to an adequate degree.

4.6. *Ablation Study*

In order to study the efficiency of the various modules in our proposed approach, we conducted an ablation study on the SigmaLaw-ABSA dataset as shown in Table 5. It is observed that removing both attention mechanisms and GCN drops the F1 score by 0.0617. Introducing the attention mechanism (with dependency tree distance) to the baseline increases the F1 score by 0.0439. This verifies the significance of the position-aware attention mechanism. The results gained from using the dependency tree distance to calculate position weights shows higher performance than the calculating position weights through basic relative distances. This verifies the impact of the syntactic information introduced by the dependency trees.

Further, we can see that the model shows higher results with the introduction of the GCN layer. Therefore we can conclude that the GCN layer contributes significantly to increase the results since it helps to capture the inter-dependencies among multiple aspects and relationships between words at long ranges.

Table 5. Results of ablation study

Setting	Accuracy	F1 score
Base	0.6287	0.5664
Base + Attention (relative distance)	0.6689	0.5989
Base + Attention (dependency tree distance)	0.6793	0.6103
Base + Attention (dependency tree distance) + 1_layer GCN	0.6938	0.6215
Base + Attention (dependency tree distance) + 2_layer GCN	0.7086	0.6281

5. Conclusion

This journal paper presents the deep learning based methodology and experiments on aspect-based sentiment analysis for the legal domain and this study is an extension of our conference paper [43]. We analysed the existing deep-learning based model architectures and pointed out suitable model components for tackling the challenges of the legal domain. Accordingly, we introduced a deep learning-based approach to perform party-based sentiment analysis in legal opinion texts. First, the model utilizes a pre-trained BERT model (further fine tuned on a legal corpus) for a strong word embedding. Then, the model employs a position-aware attention mechanism, to capture the critical parts of the sentence relevant to aspects, with incorporating position information, using the dependency tree. Because multiple legal party members are involved in a single sentence, a GCN is employed over the attention mechanism to model the inter-dependencies between members. To evaluate the proposing systems we used SigmaLaw-ABSA dataset [2] which consists of 2000 sentences taken from previous court cases. Experiments were carried out using the SigmaLaw-ABSA dataset and the experimental results demonstrate that our proposed approach outperforms all other existing state-of-art ABSA models.

References

[1] I. Rajapaksha, C. R. Mudalige *et al.*, "Rule-based approach for party-based sentiment analysis in legal opinion texts," *arXiv preprint arXiv:2011.05675*, 2020.

[2] C. R. Mudalige, D. Karunarathna *et al.*, "Sigmalaw-absa: Dataset for aspect-based sentiment analysis in legal opinion texts," *arXiv preprint arXiv:2011.06326*, 2020.

[3] C. Samarawickrama, M. de Almeida *et al.*, "Party Identification of Legal Documents using Co-reference Resolution and Named Entity Recognition," in *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*. IEEE, 2020, pp. 494–499.

[4] M. de Almeida, C. Samarawickrama *et al.*, "Legal Party Extraction from Legal Opinion Text with Sequence to Sequence Learning," in *2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer)*. IEEE, 2020, pp. 143–148.

[5] C. Samarawickrama, M. de Almeida *et al.*, "Identifying legal party members from legal opinion texts using natural language processing," EasyChair, Tech. Rep., 2021.

[6] S. Moralwar and S. Deshmukh, "Different approaches of sentiment analysis," *International Journal of Computer Sciences and Engineering*, vol. 3, no. 3, pp. 160–165, 2015.

[7] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining text data*. Springer, 2012, pp. 415–463.

[8] V. Gamage, M. Warushavithana *et al.*, "Fast Approach to Build an Automatic Sentiment Annotator for Legal Domain using Transfer Learning," in *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 2018, pp. 260–265.

[9] K. Schouten and F. Frasincar, "Survey on aspect-level sentiment analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 3, 2015.

[10] A. Bhoi and S. Joshi, "Various approaches to aspect-based sentiment analysis," *ArXiv*, vol. abs/1805.01984, 2018.

[11] M. Pontiki, D. Galanis, H. Papageorgiou *et al.*, "Semeval-2016 task 5: Aspect based sentiment analysis," 01 2016, pp. 19–30.

[12] K. Sugathadasa, B. Ayesha *et al.*, "Synergistic Union of Word2Vec and Lexicon for Domain Specific Semantic Similarity," *IEEE International Conference on Industrial and Information Systems (ICIIS)*, pp. 1–6, 2017.

[13] Supreme Court, "Lee v. United States," *US*, vol. 432, no. No. 76-5187, p. 23, 1977.

[14] V. Jayawardana, D. Lakmal *et al.*, "Deriving a Representative Vector for Ontology Classes with Instance Word Vector Embeddings," in *2017 Seventh International Conference on Innovative Computing Technology (INTECH)*. IEEE, 2017, pp. 79–84.

[15] ——, "Word Vector Embeddings and Domain Specific Semantic based Semi-Supervised Ontology Instance Population," *International Journal on Advances in ICT for Emerging Regions*, vol. 10, no. 1, p. 1, 2017.

[16] ——, "Semi-Supervised Instance Population of an Ontology using Word Vector Embedding," in *Advances in ICT for Emerging Regions (ICTer), 2017 Seventeenth International Conference on*. IEEE, September 2017, pp. 1–7.

[17] S. Jayasinghe, L. Rambukkanage, A. Silva *et al.*, "Critical sentence identification in legal cases using multi-class classification," *arXiv preprint arXiv:2111.05721*, 2021.

[18] K. Sugathadasa, B. Ayesha *et al.*, "Legal Document Retrieval using Document Vector Embeddings and Deep Learning," in *Science and Information Conference*. Springer, 2018, pp. 160–175.

[19] J. J. Nay, "Gov2vec: Learning distributed representations of institutions and their legal text," *arXiv preprint arXiv:1609.06616*, 2016.

[20] G. Ratnayaka, T. Rupasinghe *et al.*, "Shift-of-Perspective Identification Within Legal Cases," in *Proceedings of the 3rd Workshop on Automated Detection, Extraction and Analysis of Semantic Information in Legal Texts*, 2019.

[21] D. Rao and D. Ravichandran, "Semi-supervised polarity lexicon induction," in *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, 2009, pp. 675–682.

[22] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, "Target-dependent twitter sentiment classification," in *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*, 2011, pp. 151–160.

[23] X. Ding, B. Liu, and P. S. Yu, "A holistic lexicon-based approach to opinion mining," in *Proceedings of the 2008 international conference on web search and data mining*, 2008, pp. 231–240.

[24] X. Zhao, J. Jiang, H. Yan, and X. Li, "Jointly modeling aspects and opinions with a maxent-lda hybrid." ACL, 2010.

[25] R. Piryani, V. Gupta, V. K. Singh, and U. Ghose, "A linguistic rule-based approach for aspect-level sentiment analysis of movie reviews," in *Advances in Computer and Computational Sciences*. Springer, 2017, pp. 201–209.

[26] B. Plank, A. Søgaard, and Y. Goldberg, "Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss," *arXiv preprint arXiv:1604.05529*, 2016.

[27] D. Tang, B. Qin, X. Feng, and T. Liu, "Effective lstms for target-dependent sentiment classification," *arXiv preprint arXiv:1512.01100*, 2015.

[28] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based lstm for aspect-level sentiment classification," in *Proceedings of the 2016 conference on empirical methods in natural language processing*, 2016.

[29] J. Cheng, S. Zhao, J. Zhang, I. King, X. Zhang, and H. Wang, "Aspect-level sentiment classification with heat (hierarchical attention) network," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, pp. 97–106.

[30] P. Chen, Z. Sun, L. Bing, and W. Yang, "Recurrent attention network on memory for aspect sentiment analysis," in *Proceedings of the 2017 conference on EMNLP*, 2017.

[31] D. Ma, S. Li, X. Zhang, and H. Wang, "Interactive attention networks for aspect-level sentiment classification," *arXiv preprint arXiv:1709.00893*, 2017.

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

[33] P. Zhao, L. Hou, and O. Wu, "Modeling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification," *Knowledge-Based Systems*, vol. 193, p. 105443, 2020.

[34] Y. Zhang, Q. Liu, and L. Song, "Sentence-state LSTM for text representation," in *ACL*, 2018.

[35] P. Demotte, L. Senevirathne, B. Karunanayake, U. Munasinghe, and S. Ranathunga, "Sentiment analysis of sinhala news comments using sentence-state lstm networks," in *2020 Moratuwa Engineering Research Conference (MERCon)*, 2020, pp. 283–288.

[36] R. He, W. S. Lee, H. T. Ng, and D. Dahlmeier, "Effective attention modeling for aspect-level sentiment classification," in *Proceedings of the 27th international conference on computational linguistics*, 2018, pp. 1121–1131.

[37] J. Pennington, R. Socher, and C. Manning, "GloVe: Global vectors for word representation." Association for Computational Linguistics, Oct. 2014, pp. 1532–1543.

[38] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding." ACL, Jun. 2019, pp. 4171–4186.

[39] G. Ratnayaka, N. de Silva *et al.*, "Effective approach to develop a sentiment annotator for legal domain in a low resource setting," *arXiv preprint arXiv:2011.00318*, 2020.

[40] S. Gu, L. Zhang, Y. Hou, and Y. Song, "A position-aware bidirectional attention network for aspect-level sentiment analysis," in *Proceedings of the 27th International Conference on Computational Linguistics*. ACL, Aug. 2018.

[41] Q. Liu, H. Zhang, Y. Zeng, Z. Huang, and Z. Wu, "Content attention model for aspect based sentiment analysis," in *Proceedings of the 2018 World Wide Web Conference*, 2018.

[42] D. Tang, B. Qin, and T. Liu, "Aspect level sentiment classification with deep memory network," *arXiv preprint arXiv:1605.08900*, 2016.

[43] I. Rajapaksha, C. R. Mudalige *et al.*, "Sigmalaw PBSA-A Deep Learning Model for Aspect-Based Sentiment Analysis for the Legal Domain," in *International Conference on Database and Expert Systems Applications*. Springer, 2021, pp. 125–137.