Article

Semi-Supervised Model for Aspect Sentiment Detection

Zohreh Madhoushi, Abdul Razak Hamdan and Suhaila Zainudin *

Center for Artificial Intelligence Technology, Faculty of Information Science and Technology,
Universiti Kebangsaan Malaysia, Bangi 43650, Selangor, Malaysia
* Correspondence: suhaila.zainudin@ukm.edu.my

Abstract: Advancements in text representation have produced many deep language models (LMs), such as Word2Vec and recurrent-based LMs. However, there are scarce works that focus on detecting implicit sentiments with a small amount of labelled data because there are many different review areas. Deep learning techniques are suitable to automate the representation learning process. Hence, we proposed a semi-supervised aspect-based sentiment analysis (ABSA) model for online review to predict explicit and implicit sentiment in three domains (laptop, restaurant, and hotel). The datasets of this study, S1 and S2, were obtained from a standard SemEval online competition and Amazon review datasets. The proposed models outperform the previous baseline models regarding the F1-score of aspect category detection and accuracy of sentiment detection. This study finds more relevant aspects and accurate sentiment for ABSA by developing more stable and robust models. The accuracy of sentiment detection is 84.87% in the restaurant domain on the first dataset. For the second dataset, the proposed method achieved 84.43% in the laptop domain, 85.21% in the restaurant domain, and 85.57% in the hotel domain. The novelty is the proposed new semi-supervised model for aspect sentiment detection with embedded aspect inspired by the encoder–decoder architecture in the neural machine translation (NMT) model.

Keywords: aspect-based; sentiment analysis; deep learning

1. Introduction

A significant task in sentiment analysis (SA) for a product review is to process reviews and classify user opinions as positive or negative [1]. This task is doable at different levels of analysis, from the document level to the sentence and phrase level [2]. Many methods and techniques have recently been proposed for various tasks at different levels [3–7]. This study focuses on aspect-level sentiment classification. There are other groups of methods for aspect-level sentiment classification. However, regardless of the existing techniques, the most crucial point in any natural language processing (NLP) task is to find a way to make machines understand language or text. This work investigates a deep language model as the basis of a new model for aspect-level sentiment classification.

Deep learning techniques automate the process of representation learning in multi-computational layers. These techniques enabled researchers to improve the state-of-the-art for many NLP tasks, such as SA [8,9] especially, and in other domains such as image and speech. Many LMs have been developed, such as Word2Vec [10] and deep LM [11,12]. However, these emerging LMs have not yet fully addressed aspect category detection, mainly because there is no study to design experiments assessing the effect of recent advanced LMs on the specific task of aspect-level sentiment classification.

The authors of [12] explored a character-level LSTM-based LM for sentence-level sentiment classification without using labelled data. The sentiment polarity estimation methods were categorised into two machine learning methods and the Lexicon/dictionary-based method in [13]. Lexicon-based methods use sentiment lexicons, which contain a list of sentiment words to determine a given sentiment's rating [14–16]. This approach solves previous machine learning problems because there is no need for training data.

The strength of these methods is that they perform reasonably well in several areas. These methods also have weaknesses because using lexicons to find context-dependent sentiment is hard. The sentiment polarity or rating of words identified by these methods is not dependent on the reviews' context. However, some sentiments have context-dependent polarity. For example, 'quiet' is positive for a vacuum cleaner but has a negative sentiment for a speakerphone. Also, lexicon-based models can hardly detect implicit sentiments.

Classical and modern supervised models are impractical because of their supervised nature and inability to work in different areas. It is not easy to obtain performance enhancement in different supervised learning models, which means that models developed in one domain do not work in another. A considerable amount of labelled data in one domain is needed for the model to perform well. There is a lack of annotated datasets to train a model for all areas. There are many product and service review areas online, and gathering labelled data is not easy for each domain. Therefore, the method must be as domain-independent as possible.

The ABSA system characteristics are the system's ability to work in different areas without labelled data or at least with a small amount of labelled data, which means working with unlimited classes or aspects and related sentiments. It is also challenging to find the domain-dependent orientation of opinions in lexicon-based models [17,18].

For machine learning models, the same sentence-level sentiment detection methods are applied at the phrase level [19–23]. Because these models predict each sentence's sentiment, they cannot detect the sentiment of more than one aspect in a sentence. This weakness is more problematic in modern deep-learning models.

The proposed model's contributions can be summarized as follows:

- The first contribution of the current study is that a new mechanism is proposed to utilize these sentence level representations for that task of aspect category detection.
- The second contribution is that, by combining this mechanism with word-level similarity measurement, a new model for the aspect category detection is proposed.
- The final contribution of the current study is that a new semi-supervised model for aspect sentiment detection is proposed.

2. Related Works

Deep learning models are state-of-the-art for many NLP tasks because of their ability to represent a high-level feature. Recent works have studied advanced deep learning models for NLP tasks [24–27]. Most of these models are supervised models. Deep learning models need more data than classical machine learning models to learn the required features. Further, these models' incompatibility with different areas still exists because of their supervised nature. The most accurate results in the literature treat the sentiment detection of specific aspects as a classification problem and try to augment the element in the deep learning architecture to find the sentiment of particular aspects using an attention mechanism. Attentional recurrent models achieve considerable success in implicit aspect sentiment detection [25,26]. These attentional deep learning models are supervised; therefore, their performance highly depends on the amount of labelled data for the training.

Section 3 (Materials and Methods) presents the proposed aspect-embedded attentional encoder–decoder (AE-AED) model. A step-by-step explanation of the proposed architecture explaining the building blocks of the final model is included. The datasets and evaluation methods used in this study are discussed. Then, the baselines for this task are presented. This section continues with the section on Experiments and Discussion. Lastly, we conclude the paper with future works.

3. Materials and Methods

This section explains the datasets used in this study. The proposed model has two phases of unsupervised (phase 1) and semi-supervised (phase 2). The unsupervised phase is trained on a mixed dataset (M + S1 + S2) in this study's domains. S1 is the SemEval 2014 competition dataset that provided areas of laptops and restaurants [28]. The S2 dataset

is the SemEval 2015 competition dataset and provided data for laptops, restaurants, and hotels [29]. We propose that the AE-AED model that retains the sentiment of several aspects from different areas. Therefore, this study used the Amazon product review dataset M, which is a combined dataset from Amazon reviews on electronics, SemEval 2016 task 5 restaurant and laptop, Yelp restaurant reviews, and hotel reviews as a training corpus to learn the distributed sentence representation.

The aspect categories of the unsupervised phase are identified using the similarity score model. Therefore, the input of the sentiment detection model is a sentence with one specific aspect category. If the sentence has more than one aspect category, then the sentence with each aspect category is a separate input to the model. In the semi-supervised AE-AED phase, a pre-trained model of the unsupervised phase is used as the initial model to be trained more with the labeled datasets of S1 and S2.

The distribution of sentiment classes in S1 and S2 is shown in Tables 1 and 2, respectively. It is clear from the table that positive is the majority class in all areas in both S1 and S2. The sentiment distribution is imbalanced in the laptop domain in S2, while in the restaurant domain, there is a significant imbalance between the positive and negative classes across the training and test sets in S2. The same occurs for the restaurant domain positive class in S1.

Table 1. Sentiment distribution for S1 [28].

Data	Positive	Negative	Neutral
Restaurants—Train	61.46%	10.73%	21.76%
Restaurants—Test	78.95%	11.94%	21.45%

Table 2. Sentiment distribution for S2 [29].

Data	Positive	Negative	Neutral
Restaurants—Train	72.43%	24.36%	3.20%
Restaurants—Test	53.72%	40.96%	5.32%
Laptop—Train	55.87%	38.75%	5.36%
Laptop—Test	57.00%	34.66%	8.32%
Restaurants—Test	71.68%	24.77%	3.53%

We selected ten subsets from the training data S1 and S2, ranging from 10% to 100% of the training data on laptops and restaurants. No training data were available in the hotel domain in S2 to evaluate the model on less labelled data with the implicit sentiment.

In the laptop domain, 22% of the sentences in S1 and 23% of the sentences in S2 have implicit sentiment (Table 3). Similarly, in the restaurant domain, 24% of the sentences in S1 and 26% of the sentences in S2 have implicit sentiment. The subsets are selected randomly from the original training data that follow a similar proportion of unspoken sentiment and the proportion of classes for each area.

Table 3. Sentiment distribution per domain (RS—restaurants, LP—laptops), S1 and S2.

Data	Implicit Sentiment
S1-LP	22%
S1-RS	24%
S2-LP	23%
S2-RS	26%

3.1. Proposed Aspect Sentiment Classification Model

We proposed a model that addresses aspect sentiment detection tasks using LSTM, attention mechanism, and encoder–decoder architecture with embedded aspect. The model

is called aspect-embedded attentional encoder–decoder (AE-AED). The model is based on the encoder–decoder LM with attention, which needs a small amount of labelled data. The encoder part of this LM is trained on new data with the same architecture. The idea is to train the attentional decoder part of the model with the new dataset, aspect augmentation, and SoftMax classifier.

The decoder part of the same model is trained with new pre-processed data to classify the sentiment of aspects for multi-domain online reviews. The AE-AED model has an unsupervised part followed by a semi-supervised part and classifier with labelled data for each domain.

3.2. Sentiment Detection Model

The idea of detecting sentiment without many labelled data is to find a good representation of a sentence, which is well enough to represent a sentence's sentiment concerning the specific aspect. This study tries to determine if vector representations of the sentences, generated by the LM, contain each aspect's sentiment information. If so, then with a small amount of labelled data from each domain, the model can be trained to predict sentiment regarding specific aspects. The proposed model has two phases—unsupervised (phase 1) and semi-supervised (phase 2). The unsupervised phase is trained on a mixed dataset (M + S1 + S2) in this study's domains. The aspect categories of the unsupervised phase are identified from the similarity score model. The process is presented in Figure 1. Therefore, the input of the sentiment detection model is a sentence with one specific aspect category. If the sentence has more than one aspect category, then each aspect category of the sentence is a separate input to the model. Therefore, the model detects sentiment for all of the detected aspect categories for the sentence. In the semi-supervised AE-AED phase, a pre-trained model of the unsupervised phase is used as the initial model to be trained more with the labelled datasets of S1 and S2. Considering that the objective is to develop a model that needs a small amount of labelled data compared with the state-of-the-art models, in this study, an experiment is conducted to find the amount of data required for this model compared with the state-of-the-art models.

The input of this process is a sentence list with aspect categories. It is necessary to pre-process the data and repeat each sentence for each aspect category in a separate line. One sentence relates to one aspect per line. For S1 and S2, the sentence list with aspect categories presented in the data is pre-processed. The sentence list with aspect categories is pre-processed for dataset M.

The following section explains the encoder–decoder LSTM. Then, the proposed model is presented. The proposed model is an encoder–attention–decoder LSTM with an embedded aspect followed by a classifier. As the aspect is embedded in our proposed model, the new model is called the aspect-embedded attentional encoder–decoder (AE-AED). The following sections further describe the building blocks of AE-AED.

3.3. Attentional LSTM

Adding an attention layer to LSTM helps the network to capture the key part of the sentence for a given aspect. The attention mechanism will produce an attention weight vector α and a weighted hidden representation r. Let $H \in R^{d \times N}$ be a matrix consisting of hidden vectors $[h_1, \ldots, h_N]$ produced by the LSTM, where d is the size of the hidden layers and N is the length of the given sentence. Furthermore, v_a represents the embedding of the aspect and $eN \in R^N$ is a vector of 1 s. r is computed as follows:

$$\alpha = softmax \left(w^T M \right) \tag{1}$$

$$r = H\alpha^T \tag{2}$$

where $M \in R^{(d+da) \times N}$, $\alpha \in R^N$, $r \in R^d$. $W_h \in R^{d \times d}$, $W_v \in R^{da \times da}$, and $w \in R^{d+da}$ are projection parameters. α is a vector consisting of attention weights and r is a weighted representation of sentence with a given aspect. The operator (a circle with a multiplication

sign inside) means $v_a \otimes eN = [v; v; \ldots ; v]$; that is, the operator repeatedly concatenates v for N times, where eN is a column vector with N 1 s. $W_v v_a \otimes eN$ repeats the linearly transformed v_a as many times as there are words in the sentence. The final sentence representation is given by the following:

$$h^* = tanh\,(W_p r + W_x h_N) \tag{3}$$

where $h^* \in R^d$, W_p, and W_x are projection parameters to be learned during training. Figure 2 illustrates the attentional BLSTM architecture.
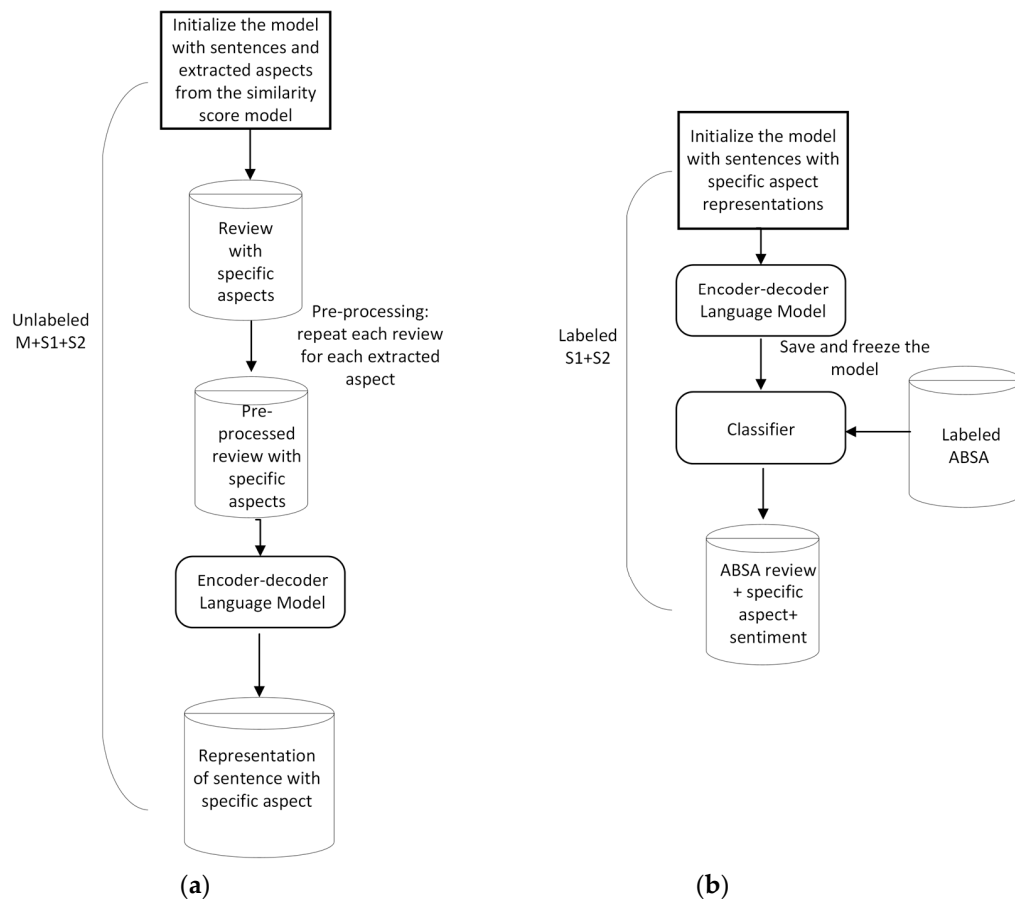

(a) **(b)**

Figure 1. Sentiment detection process: phase 1 (**a**) and phase 2 (**b**).
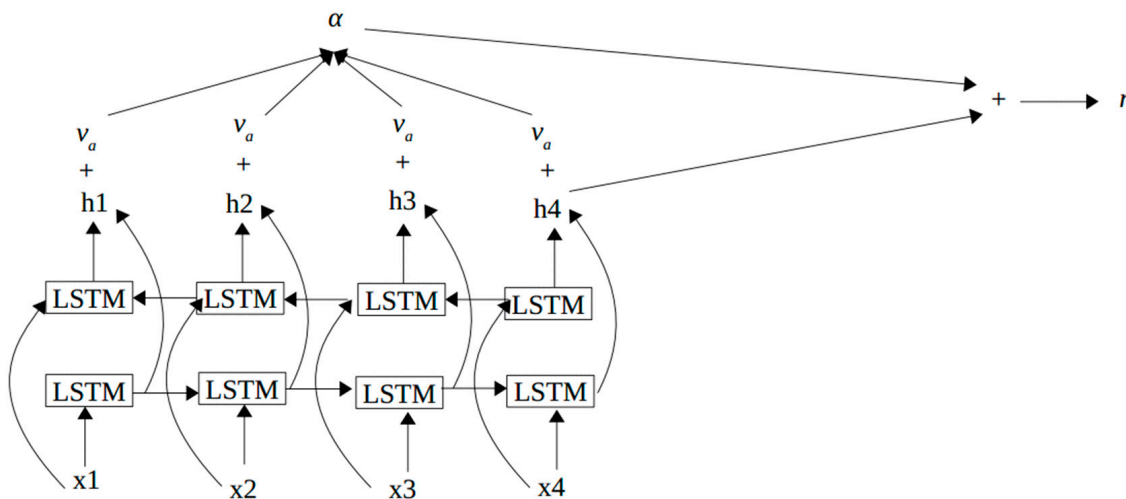

Figure 2. Attentional BLSTM architecture.

3.4. Encoder–Decoder Model

The encoder and decoder choice of this model can be any type of RNN such as GRU and LSTM. It consists of an encoder for a source language and a decoder for a target language. The idea is that all RNNs can be trained to map an input sequence to an output sequence. The encoder RNN obtains the input sequence and produces the context c, which is usually the final hidden states of the RNN. The decoder is often trained to predict the next word y_t, given the context vector c and all previously predicted words $\{y_1, \dots, y_{t-1}\}$. In other words, the decoder defines a probability over the translation y by decomposing the joint probability into the ordered conditionals:

$$p(Y) = \prod_{t=1}^{T} p(\{y_1, \dots, y_{t-1}\}, c) \tag{4}$$

where y = $(y_1, \dots, y_{Ty})$. With an RNN, each conditional probability is modeled as follows:

$$p(y_t | \{y_1, \dots, y_{t-1}\}, c) = g(y_t | y_{t-1}, s_t, c) \tag{5}$$

where g is a nonlinear, potentially multi-layered function that outputs the probability of y_t, and s_t is the hidden state of the RNN. The limitation of this architecture is that the context vector c cannot properly summarize a long sequence. This problem is detected by [30]. Hence, they solved the issue using attentional mechanism explained in the previous section. Using attention weights; each decoder output depends on a weighted combination of all of the input states, not just the last state. Figure 3 shows the architecture of an attentional encoder–decoder depicted from [30].
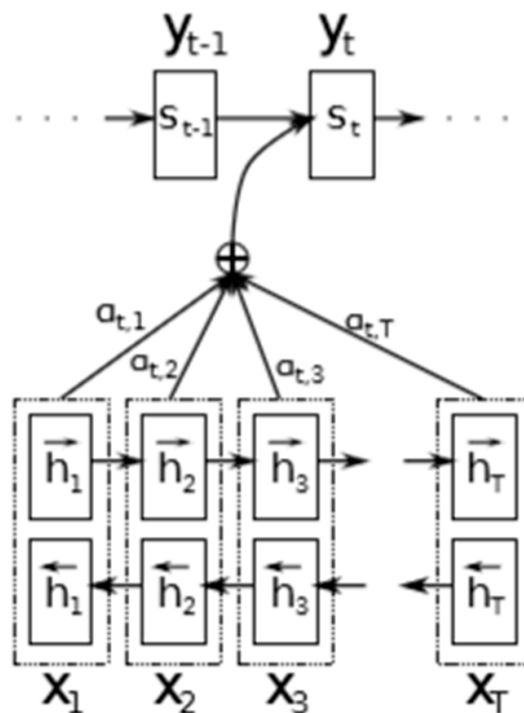

Figure 3. Attentional encoder (bidirectional) decoder [30].

Next, the neural translation model is described. The encoder components are denoted by index j, and the referred decoder components by index i. The same annotation is followed in this work. At each time-step i, the attention mechanism computes a different context vector ci $\in$ R^{2H} as the weighted sum of the sequence of annotations h_1^j:

$$c_i = \sum_{j=1}^{j} \alpha_{ij} h_j \tag{6}$$

where $\alpha_{ij} \in$ R is the weight assigned to each annotation h_j. This weight is computed by means of the SoftMax function:

$$\alpha_{ij} = \frac{\exp(a_{ij})}{\sum_{k=1}^{j} \exp(a_{ij})} \tag{7}$$

where $a_{ij} \in R$ is a score provided by a soft alignment model, which measures how well the inputs from the source position j and the outputs around the target position i match. This alignment model is implemented by a perceptron with N units:

$$a_{ij} = V_a^T \tan h \left(W_a s_{i-1} + U_a h_j \right) \tag{8}$$

where $s_{i-1} \in R^H$ is the hidden state from the decoder; $\tanh(\bullet)$ is applied element-wise; and $v_a \in R^N$, $W_a \in R^{N \times H}$, and $U_a \in R^{N \times 2H}$ are the weight matrices. The decoder is an RNN with GRU units, which generates the translated sentence $y_1^I = y_1, \dots, y_I$. Each word y_i depends on the previously generated word y_{i-1}, the current hidden state of the decoder si, and the context vector ci; the probability of a word at the time-step i is defined as follows:

$$p(y_i | y_1^{i-1}, x_1^j; \theta\}) = \bar{y}_i^T \varphi(V\eta | y_{i-1}, s_i, c_i) \tag{9}$$

where $\varphi(\cdot) \in R^{|Vy|}$ is a SoftMax function that produces a vector of probabilities, $|Vy|$ is the size of the target vocabulary, $\bar{y}_i \in N^{|Vy|}$ is the one-hot representation of the word y_i, $V \in R^{|Vy| \times L}$ is the weight matrix, and η is the output of LSTM units with an L-sized maxout output layer.

3.5. Aspect-Embedded Attentional Encoder–Decoder (AE-AED) Model

In this study, the aspect category is embedded in an attentional NMT to propose AE-AED. Putting all of the above together, the final architecture is very similar to NMT [30], as shown in Figure 3. However, the goal in AE-AED is to find a new representation for sentences that focuses on a specific aspect, whereas the goal in NMT is to translate a sentence from one language to another language. Therefore, in AE-AED, both the source and target languages are the same (English language). Another difference between AE-AED and NMT is that the related aspect of a sentence is embedded in the AE-AED model to obtain the aspect-specific representation of a sentence. Before the AE-AED model is trained on S1 and S2, the encoder–decoder model without aspect augmentation is trained using dataset M and unlabelled S1 and S2. The objective is to predict the next word. After training the encoder–decoder LM, the encoder part of the model is frozen, then the attentional decoder with aspect augmentation is trained on M, S1, and S2. The same encoder representation for a list of pre-known aspects is used. The hidden layer of this encoder is concatenated with the hidden layer of encoder in a neural model translation with attention, shown in Figure 3. The model does not use any new encoder to obtain the aspect representations. Therefore, for the AE-AED model, Equation (6) is changed to the following:

$$c_i = \sum_{j=1}^{j} \alpha_{ij} \left(h_j + ha_j \right) \tag{10}$$

The final output layer was a three-dimensional SoftMax layer representing each output class. SoftMax classifier is used on top of the new sentence representations on 10% of dataset S1 and S2 for polarity prediction (Figure 4).

We compare our proposed approach (AE-AED) with several deep learning and none-deep learning models on S1 and S2. The deep learning models include Bidirectional form of target dependent LSTM which is called TD-LSTM [31] and name it as TD-BLSTM. The second baseline is [24] which integrates CRF with Recursive Neural Network and add linguistic features. The third baseline is [25] which applied attention with bidirectional GRU model to attend the aspect information for one given aspect and extract sentiment for that given aspect. The final recent baseline is [26]. They proposed an LSTM base model which combines implicit and explicit knowledge. The model adopted a sequence-encoder and a self-attention mechanism to calculate and incorporate common-sense knowledge into LSTM-based model to jointly extract aspect categories and predict sentiment for them. A non-deep learning baseline is the winner of SemEval 2015 competition on S2 dataset which is supervised. The best accuracy for this dataset were achieved by Sentiue [32]

with a Maximum Entropy classifier along with features based on n-grams, POS tagging, lemmatization, negation words and publicly available sentiment lexica (MPQA, Bing Liu's lexicon, AFINN) for laptop and restaurant domain. Another non-deep learning baseline is the winner of SemEval 2014, NRC-Ca [33] on S1. The model uses SVM along with several lexical features.
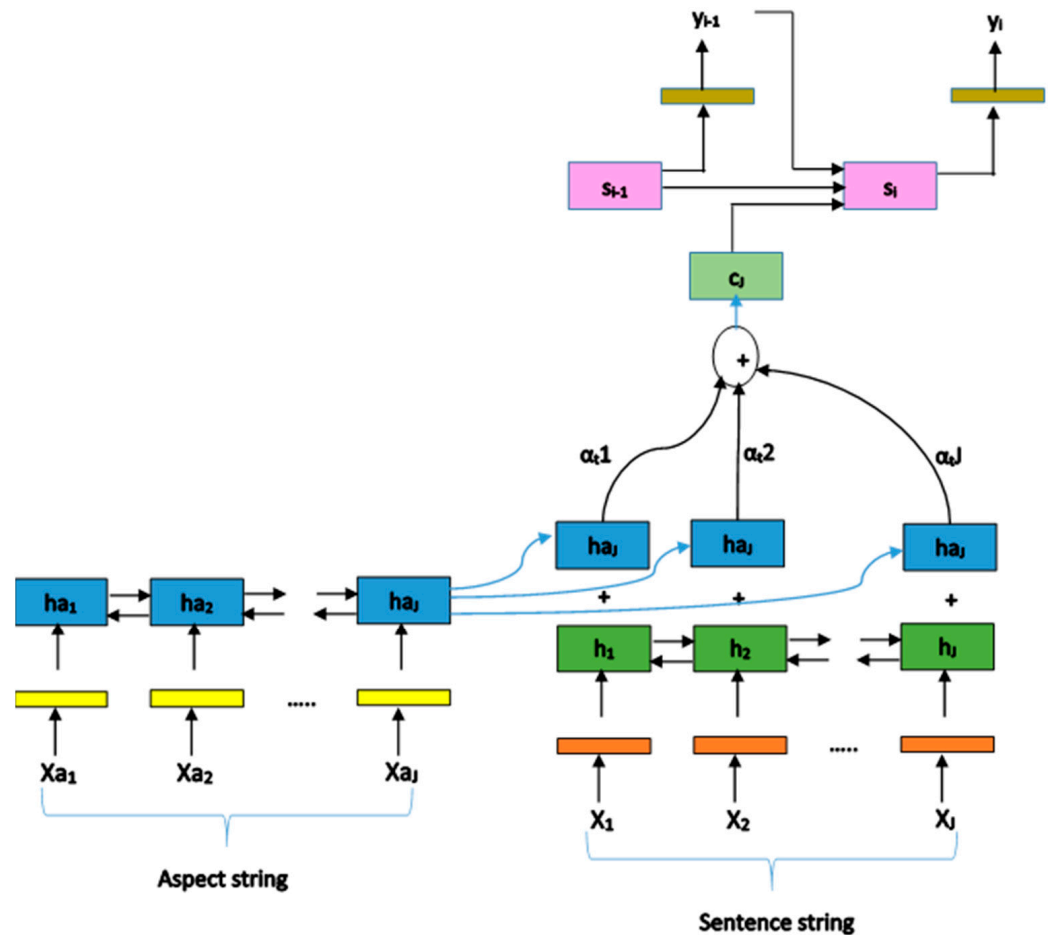

Figure 4. AE-AED model architecture.

The last non-deep learning baseline is the unsupervised baseline of V3 [17] on S1 dataset. They have used the SentiWords of [34] and lexicon of [35] as a sentiment lexicons. Using direct dependency relations between aspect terms and sentiment bearing words they assign the sentiment value from the lexicon to the aspect term. They make a simple count of the sentiments of the aspect terms classified under a certain category to assign the sentiment of that category in a particular sentence.

3.6. Model Selection

The accuracy of the two-layer bi-directional LSTM (Bi-LSTM-2L) and two-layer bi-directional GRU (Bi-GRU-2L) is evaluated on the sentiment detection task on the final model. The result is shown in Table 4. As the result shows the LSTM base model is more accurate. Therefore, the two-layer bi-directional LSTM is used for both the encoder and decoder of the model on all areas of the laptop and restaurant domains on the S2 dataset.

Table 4. Sentiment detection accuracy results for the two architectures.

Model/Domain	Laptop	Restaurant	Hotel
Bi-LSTM-2L	84.43%	85.21%	83.93%
Bi-GRU-2L	83.37%	83.58%	80.49%

Continued training can be sensitive to the learning rate. Therefore, this study runs a continued training experiment over four learning rates (0.1, 0.25, 0.50, and 0.75) and chooses the best result based on the average accuracy on the test set. These learning rates are the most common learning rates to test for the encoder–decoder architecture. As is clear in Figure 5, the best result is with a learning rate of 0.5.
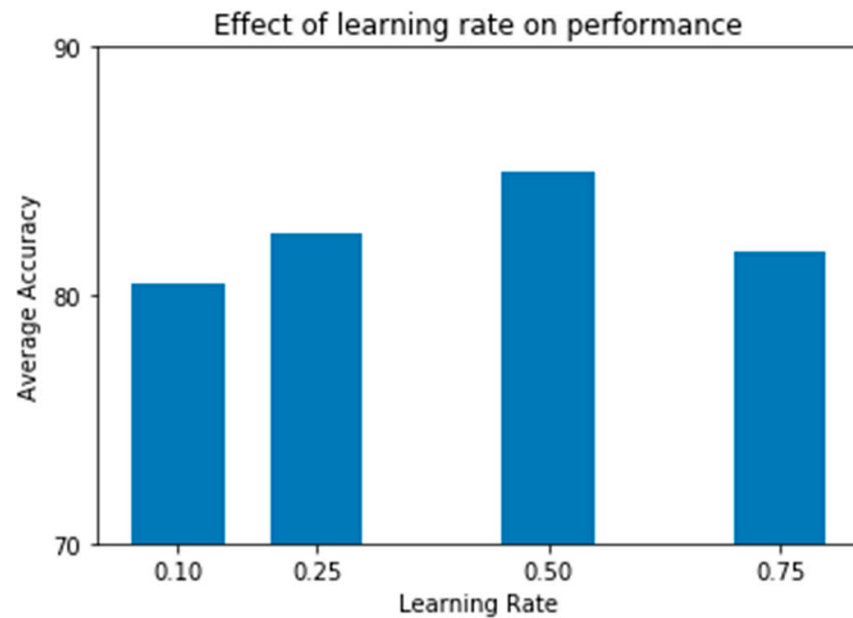

Figure 5. Effect of learning rate on performance.

To investigate how the accuracy changes by increasing the number of records in the training data, the accuracy of the best model for 10 subsets of training data is analyzed. The trend is shown in Figures 6–8 for S1 and S2.
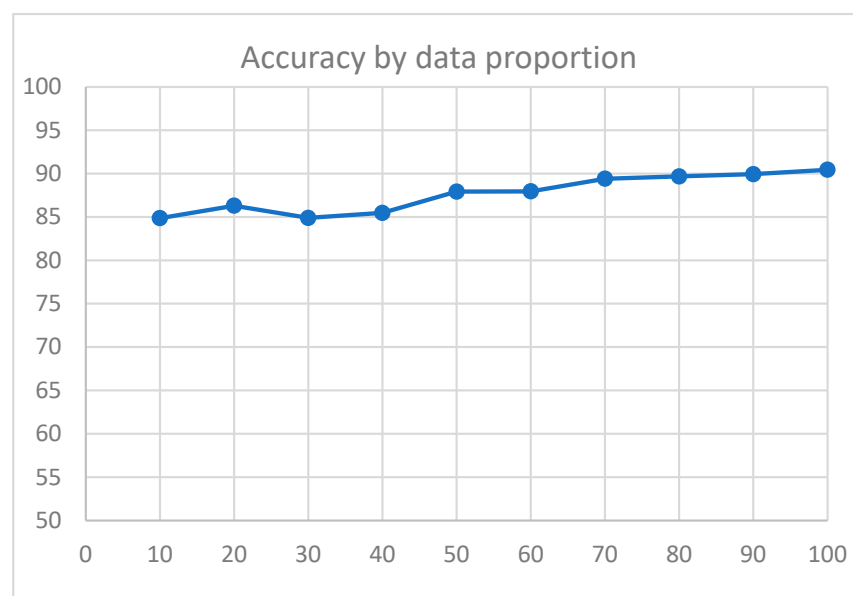

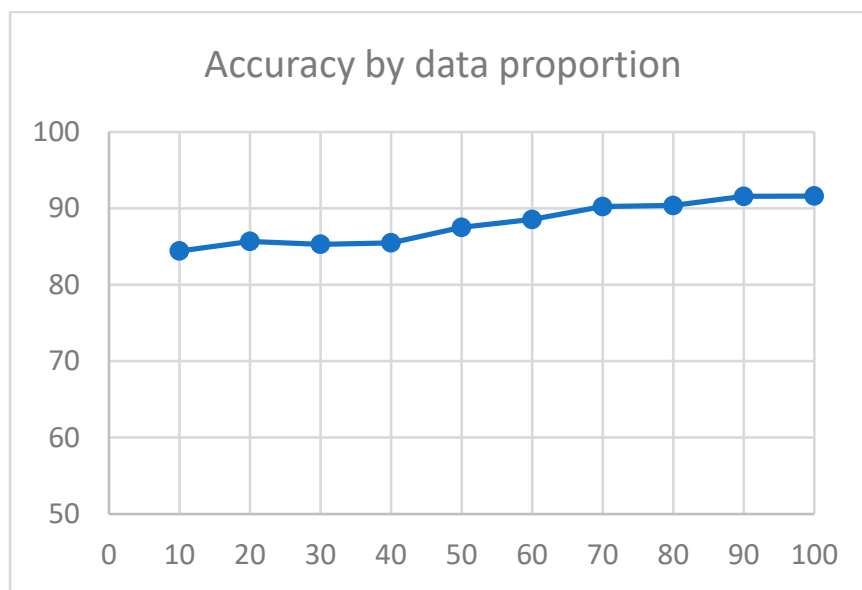**Figure 6.** Accuracy of sentiment detection by increasing the data proportion on restaurant S1.

Figure 7. Accuracy of sentiment detection by increasing the data proportion on laptop S2.
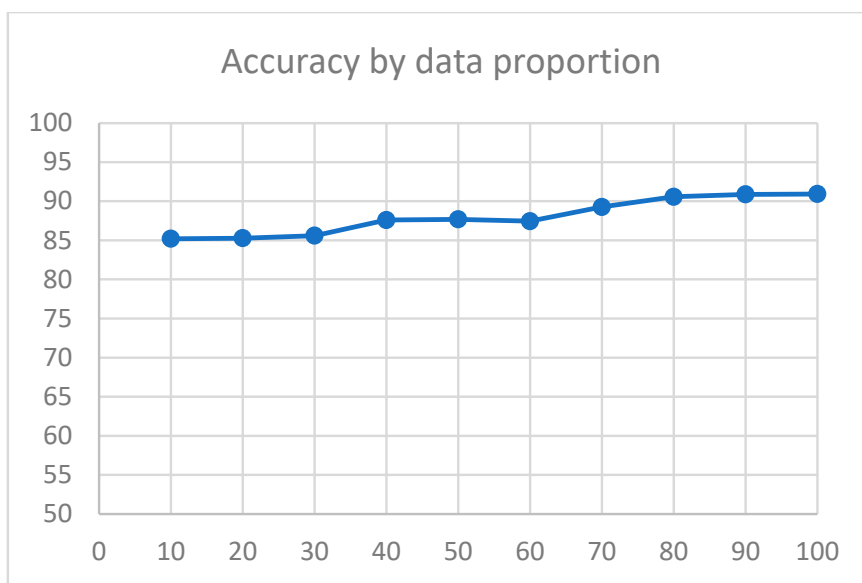
Figure 8. Accuracy of sentiment detection by increasing the data proportion on restaurant S2.

4. Results

To evaluate the model on less labelled data with implicit sentiment, which is the third objective of this study, 10 subsets were selected from the training data, ranging from 10% to 100% of the training data on laptop and restaurant. There were no training data available in hotel area in S2. In the laptop area, 22% of the sentences in S1 and 23% of the sentences in S2 has implicit sentiment, respectively, as shown in Table 5. Similarly, in the restaurant area, 24% of the sentences in S1 and 26% of the sentences in S2 had implicit sentiment, respectively, as shown in Table 5. The subsets were selected randomly from the original training data, which follow the similar proportion of implicit sentiment and similar proportions of classes for each area.

Word embedding is vector representation for words. The commonly used ones are random initialization and unsupervised pre-training of word embedding. Our experiment used unsupervised pre-training of the Word2Vec method on dataset A. Then, the word vectors were fine-tuned along with other model parameters during training.

Table 5. Sentiment distribution per domain (RS—restaurants, LP—laptops) for S1 [28] and S2 [29].

Data	Implicit Sentiment
S1-LP	22%
S1-RS	24%
S2-LP	23%
S2-RS	26%

Vectors for all training sentences were extracted from the encoder part of the AE-AED. Then, the decoder part of AE-AED was trained with aspects on the related domain (dataset M). The model was then frozen for the second time and a classifier was added on top of the decoder's extracted representations, with no additional fine-tuning or backpropagation through the encoder and decoder part of the model. The result of sentiment detection is shown in Tables 6 and 7.

Table 6. Sentiment detection accuracy on S1 (restaurant).

Reference	Model Name	Accuracy
[17]	V3	47.21%
[33]	NRC-Can	82.92%
[25]	Hierarchical Attention	85.10%
Proposed approach	AE-AED	84.87%

Table 7. Sentiment detection accuracy on S2.

Reference	Model Name	Laptop	Restaurant	Hotel
[17]	V3	68.38%	69.46%	71.09%
[32]	Sentiue	79.34%	78.69%	85.84%
[31]	TD-BLSTM	82.7%	-	-
[24]	Recursive Neural Conditional Random Fields	79.44%	84.14%	-
[25]	Hierarchical Attention	85.11%	80.50%	-
[26]	Joint Aspect Sentiment Model	-	74.11%	-
Proposed approach	AE-AED	84.43%	85.21%	85.57%

The lexicon-based model V3 shows an inferior performance on S1 compared with AE-AED. The dataset has implicit sentiments and sentiment keywords, not just adjectives. Therefore, a simple dependency relation cannot extract the right sentiment words. The result of V3 is better on S2, but still significantly lower compared with the AE-AED model on both datasets S1 and S2.

The AE-AED model results are better than the non-deep learning baseline on S1 (NRC-Can) [33] and the non-deep learning baseline on S2 (Sentiue) [32] (Table 7). This result shows that automatically extracted features in deep learning models can be better than hand-engineered features in classical machine learning for sentiment detection.

5. Discussion

The results show that AE-AED is comparable to deep learning baselines and TD-BLSTM, where representations are learned directly for the specific task at hand on a complete label data. This indicates that an encoder–decoder with attention to a specific aspect category can extract the feature representation for sentiment detection of one particular aspect category from an extensive related dataset. Therefore, using only 10% of the labelled data, this study competed with deep learning rivals on fully labelled S1 and S2. One drawback of fully supervised deep learning models is that they rely on the representations they obtain from labelled datasets S1 and S2. The strength of deep learning models is from the features they extract from large datasets; therefore, the performance of aspect sentiment

representation removed from only S1 or S2 relies on the reviews used in these datasets to classify the correct sentiment related to a specific aspect. Using these models on a new dataset and domain makes it very hard.

In contrast, the AE-AED model is trained on an extensive dataset in a related domain (M) to understand sentiment representations. The result shows that unsupervised pre-training helps to increase the robustness of models by seeing more variation of the same aspects in an extensive dataset compared with S1 and S2. Thus, more precise sentiments are detected. The AE-AED's strength is that it works only with 10% of the labelled S1 and S2 and achieves comparable results with the baselines. Although no labelled data are available for the hotel domain in S2, AE-AED classifies the hotel review sentences with 85.57% accuracy. The presentations learnt on the hotel review area in the unsupervised phase helped the model classify sentences in this area without any labelled data. The results of [25] are slightly better in the restaurant area on S1 and the laptop domain on S2. However, the AE-AED results are more stable in all three areas, restaurants, laptops, and hotels, across different datasets of S1 and S2. It is also clear from the results that AE-AED is more stable on separate datasets and areas than all other baselines in this study.

This study proposed a model for online review sentiment detection that finds more accurate sentiment with a few labelled data for the multi-domain dataset. Based on the result, our model works better than most of the baselines that use fully labelled data and works in three different areas mainly because the representations for specific aspect sentiments are generated from deep LM trained on a related domain.

6. Future Works and Conclusions

The first suggestion for future research is to interpret the results to provide a transparent view of the developed model that researchers can use to improve the results even further. It is not easy to solve these complex neural models. All deep models show little transparency concerning their inner workings. As a result of a complicated procedure, a typical model often lacks a reasonable explanation or understanding of its computation. This shortcoming could be problematic for developing new methods for real-world applications. For example, researchers need to understand the hidden layer's results to extend and improve the practices. Besides, ordinary users often require justifications for the model's prediction. Interpreting the proposed model results using visualisation techniques can be another right direction for future research. Unsupervised DL models such as restricted Boltzmann machines (RBMs) and more recent DL models such as generative adversarial networks (GANs) are more recent architectures applied to NLP tasks. These models are unsupervised and do not need labelled data. Another direction for future research is to work on these architectures instead of the encoder–decoder architecture.

The proposed semi-supervised sentiment detection model, AE-AED, works in three domains. In the implementation details, the Word2Vec training parameters are presented, and the best parameters are decided based on the result of the sentiment detection task at the review level. The encoder–decoder model selection and best learning rates of the AE-AED model are discussed. The best result was a learning rate of 0.5. The result is shown in terms of sentiment detection accuracy on this study's datasets and compared with the baselines. The result is presented on ten random portions of the datasets to test the performance of the AE-AED by increasing the labelled volume. An equal ratio of classes for each area is comparable to or higher than the baseline models on the completely labelled datasets.

Deep learning models have shown considerable improvement in ABSA and all other NLP tasks. However, these models are domain-dependent and need many labelled data in different domains. This study proposed a new model based on deep learning architectures for each task of ABSA. This study shows that the new models are experimentally better or at least comparable to the benchmark models.

To conclude, we proposed AE-AED for ABSA tasks using deep learning architectures, which solves the problems identified in the benchmarks. The result of this study shows the

power of unsupervised neural LM and fine-tuning it on in-domain data for both tasks of this study, aspect category detection and sentiment detection. These results can also be used in other computer systems, such as recommendation systems, to explain recommendations. The AE-AED can place a product ad with similar rated aspects in an advertising system. Other possible benefactors are business tasks related to sale management, reputation management, and public relations.

References

1. Madhoushi, Z.; Hamdan, A.R.; Zainudin, S. Sentiment analysis techniques in recent works. In Proceedings of the 2015 Science and Information Conference (SAI), London, UK, 28–30 July 2015; IEEE: Piscataway, NJ, USA.
2. Madhoushi, Z.M.Z.; Hamdan, A.R.; Zainudin, S. Aspect-Based Sentiment Analysis Methods in Recent Years. *Asia Pac. J. Inf. Technol. Multimed.* **2019**, *8*, 79–96. [CrossRef]
3. Ibrahim, S.A.; Bakar, A.A.; Yaakub, M.R.; Darwich, M. Beyond Sentiment Classification: A Novel Approach for Utilizing Social Media Data for Business Intelligence. *Int. J. Adv. Comput. Sci. Appl. IJACSA* **2020**, *11*, 437–441.
4. Awwalu, J.; Abu Bakar, A.; Yaakub, M.R. Hybrid N-gram model using Naïve Bayes for classification of political sentiments on Twitter. *Neural Comput. Appl.* **2019**, *31*, 9207–9220. [CrossRef]
5. Al-Ghuribi, S.M.; Noah, S.A.M.; Tiun, S. Unsupervised Semantic Approach of Aspect-Based Sentiment Analysis for LargeScale User Reviews. *IEEE Access* **2020**, *8*, 218592–218613. [CrossRef]
6. Adel, H.; Dahou, A.; Mabrouk, A.; Elaziz, M.A.; Kayed, M.; El-Henawy, I.M.; Alshathri, S.; Ali, A.A. Improving Crisis Events Detection Using DistilBERT with Hunger Games Search Algorithm. *Mathematics* **2022**, *10*, 447. [CrossRef]
7. Chennafi, M.E.; Bedlaoui, H.; Dahou, A.; Al-Qaness, M.A.A. Arabic Aspect-Based Sentiment Classification Using Seq2Seq Dia-lect Normalization and Transformers. *Knowledge* **2022**, *2*, 388–401. [CrossRef]
8. Sachan, D.S.; Zaheer, M.; Salakhutdinov, R. Revisiting LSTM Networks for Semi-Supervised Text Classification via Mixed Objective Function. *Proc. AAAI Conf. Artif. Intell.* **2019**, *33*, 6940–6948. [CrossRef]
9. Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; Liu, P.J. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.* **2020**, *21*, 5485–5551.
10. Li, Y.; Pang, X.; Pang, M. Adversarial Attacks on Word2vec and Neural Network. In Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence, Sanya, China, 21–23 December 2018.
11. Gu, Y.; Gu, M.; Long, Y.; Xu, G.; Yang, Z.; Zhou, J.; Qu, W. An enhanced short text categorization model with deep abundant representation. *World Wide Web* **2018**, *21*, 1705–1719. [CrossRef]
12. Alec Radford, R.J. Ilya Sutskever, Learning to Generate Reviews and Discovering Sentiment. *arXiv* **2017**, arXiv:1704.01444.
13. Truşcă, M.M.; Wassenberg, D.; Frasincar, F.; Dekker, R. A Hybrid Approach for Aspect-Based Sentiment Analysis using Deep Contextual Word Embeddings and Hierarchical Attention. In Proceedings of the Web Engineering: 20th International Conference, ICWE 2020, Helsinki, Finland, 9–12 June 2020; pp. 365–380. [CrossRef]
14. Lal, M.; Asnani, K. Aspect Extraction & Segmentation in Opinion Mining. *Int. J. Eng. Comput. Sci.* **2017**, *3*. Available online: http://www.ijecs.in/index.php/ijecs/article/view/461 (accessed on 1 January 2023).
15. Marrese-Taylor, E.; Matsuo, Y. Replication issues in syntax-based aspect extraction for opinion mining. *arXiv* **2017**, arXiv:1701.01565.
16. Nguyen, H.T.; Vo, Q.H.; Nguyen, M.L. A Deep Learning Study of Aspect Similarity Recognition. In Proceedings of the 2018 10th International Conference on Knowledge and Systems Engineering (KSE), Ho Chi Minh City, Vietnam, 1–3 November 2018.
17. Pablos, A.G.; Cuadros, M.; Rigau, G. V3: Unsupervised Generation of Domain Aspect Terms for Aspect Based Sentiment Analysis. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland, 23–24 August 2014; pp. 833–837. [CrossRef]

18. Poria, S.; Chaturvedi, I.; Cambria, E.; Bisio, F. Sentic LDA: Improving on LDA with semantic similarity for aspect-based sentiment analysis. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 4465–4473.

19. Blair-Goldensohn, S.; Hannan, K.; McDonald, R.; Neylon, T. Building a Sentiment Summarizer for Local Service Reviews. In Proceedings of the WWW 2008 Workshop: NLP in the Information Explosion Era (NLPIX 2008), Beijing, China, 22 April 2008.

20. De Albornoz, J.C.; Plaza, L.; Gervás, P.; Díaz, A. A Joint Model of Feature Mining and Sentiment Analysis for Product Review Rating. In *Advances in Information Retrieval*; Springer: Berlin/Heidelberg, Germany, 2011.

21. Tan, C.; Lee, L.; Tang, J.; Jiang, L.; Zhou, M.; Li, P. User-level sentiment analysis incorporating social networks. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 21–24 August 2011; pp. 1397–1405. [CrossRef]

22. Wei, W.; Gulla, J. Sentiment Learning on Product Reviews via Sentiment Ontology Tree. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden, 11–16 July 2010; pp. 404–413.

23. Appel, O.; Chiclana, F.; Carter, J.; Fujita, H. Cross-ratio uninorms as an effective aggregation mechanism in sentiment analysis. *Knowl. Based Syst.* **2017**, *124*, 16–22. [CrossRef]

24. Wang, W.; Pan, S.J.; Dahlmeier, D.; Xiao, X. Recursive Neural Conditional Random Fields for Aspect-based Sentiment Analysis. *arXiv* **2016**, arXiv:1603.06679. [CrossRef]

25. Cheng, J.; Zhao, S.; Zhang, J.; King, I.; Zhang, X.; Wang, H. Aspect-level Sentiment Classification with HEAT (HiErarchical ATtention) Network. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, Singapore, 6–10 November 2017; pp. 97–106. [CrossRef]

26. Ma, Y.; Peng, H.; Khan, T.; Cambria, E.; Hussain, A. Sentic LSTM: A Hybrid Network for Targeted Aspect-Based Sentiment Analysis. *Cogn. Comput.* **2018**, *10*, 639–650. [CrossRef]

27. Fan, F.; Feng, Y.; Zhao, D. Multi-grained Attention Network for Aspect-Level Sentiment Classification. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018. [CrossRef]

28. Pontiki, M.; Galanis, D.; Pavlopoulos, J.; Papageorgiou, H.; Androutsopoulos, I. Suresh ManandharSemEval-2014 Task 4: Aspect Based Sentiment Analysis. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland, 23–24 August 2014.

29. Pontiki, M.; Galanis, D.; Papageorgiou, H.; Manandhar, S.; Androutsopoulos, I. SemEval-2015 Task 12: Aspect Based Sentiment Analysis. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, CO, USA, 4–5 June 2015; Association for Computational Linguistics: Stroudsburg, PA, USA.

30. Bahdanau, D.; Cho, K.; Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv* **2014**, arXiv:1409.0473.

31. Tang, D.; Qin, B.; Feng, X.; Liu, T. Effective LSTMs for Target-Dependent Sentiment Classification. *arXiv* **2015**, arXiv:1512.01100.

32. Saias, J. *Sentiue: Target and Aspect based Sentiment Analysis in SemEval-2015 Task 12*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2015. [CrossRef]

33. Kiritchenko, S.; Zhu, X.; Cherry, C.; Mohammad, S. NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland, 23–24 August 2014. [CrossRef]

34. Guerini, M.; Gatti, L.; Turchi, M. Sentiment Analysis: How to Derive Prior Polarities from SentiWordNet. *arXiv* **2013**, arXiv:1309.5843.

35. Warriner, A.B.; Kuperman, V.; Brysbaert, M. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behav. Res. Methods* **2013**, *45*, 1191–1207. [CrossRef] [PubMed]