

Article

Research on Aspect-Level Sentiment Analysis Based on Text Comments

Jing Tian ¹, Wushour Slamu ^{2,*}, Miaomiao Xu ³, Chunbo Xu ³ and Xue Wang ⁴

- ¹ Xinjiang Multilingual Information Technology Laboratory, Xinjiang Multilingual Information Technology Research Center, College of Software, Xinjiang University, Urumqi 830017, China; tjing@stu.xju.edu.cn
- ² Xinjiang Multilingual Information Technology Laboratory, Xinjiang Multilingual Information Technology Research Center, College of Information Science and Engineering, Xinjiang University, Urumqi 830017, China
- ³ College of Information Science and Engineering, Xinjiang University, Urumqi 830017, China; xmm@stu.xju.edu.cn (M.X.); xuchunboc@gmail.com (C.X.)
- ⁴ College of Software, Xinjiang University, Urumqi 830017, China; xue15235@gmail.com
- * Correspondence: wushour@xju.edu.cn

Abstract: Sentiment analysis is the processing of textual data and giving positive or negative opinions to sentences. In the ABSA dataset, most sentences contain one aspect of sentiment polarity, or sentences of one aspect have multiple identical sentiment polarities, which weakens the sentiment polarity of the ABSA dataset. Therefore, this paper uses the SemEval 14 Restaurant Review dataset, in which each document is symmetrically divided into individual sentences, and two versions of the datasets ATSA and ACSA are created. ATSA: Aspect Term Sentiment Analysis Dataset. ACSA: Aspect Category Sentiment Analysis Dataset. In order to symmetrically simulate the complex relationship between aspect contexts and accurately extract the polarity of emotional features, this paper combines the latest development trend of NLP, combines capsule network and BERT, and proposes the baseline model CapsNet-BERT. The experimental results verify the effectiveness of the model.

Keywords: aspect-based sentiment analysis; SemEval 14 Restaurant Review dataset; BERT; capsule network



Citation: Tian, J.; Slamu, W.; Xu, M.; Xu, C.; Wang, X. Research on Aspect-Level Sentiment Analysis Based on Text Comments. *Symmetry* **2022**, *14*, 1072. <https://doi.org/10.3390/sym14051072>

Academic Editor: László T. Kóczy

Received: 3 March 2022

Accepted: 20 May 2022

Published: 23 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Sentiment Analysis refers to the process of analyzing, processing, and extracting subjective text with emotion by using natural language processing and text mining techniques. Currently, text sentiment analysis research covers a wide range of fields including natural language processing, text mining, information retrieval, information extraction, and machine learning, and has gained the attention of many scholars and research institutions. Aspect-level sentiment analysis (ABSA), a text analysis technique, can classify data by different aspects and identify the sentiment polarity of each aspect sentence [1]. Text analytics technology correlates specific sentiments with different aspects of a product or service, and further identifies the strengths and weaknesses of the product or service through user reviews, and addresses them well. ABSA obtains the reviewer's sentiment tendency towards an attribute by describing the attribute of an entity [2,3]. For example, in the sentence: Dinner was ok, service was so-so, Dinner and service are two attributes. Their emotional polarities are one positive and one negative.

Earlier, aspect-level sentiment analysis extracted features from single sentence comment attitudes, and three sentences were used as independent subtasks of sentiment analysis [4], and these three sentence-level sentiment analysis subtasks usually used the "BIO" and "BIEOS" annotation patterns based on the deep learning algorithm model [5], where B, I, E, S, O represent the starting position, internal position, ending position, single word, and non-evaluation object opinion position of sentiment polarity, respectively. Most of the researchers designed three independent models to analyze the research objects,

resulting in the inability of the three tasks to correlate with each other and to establish an effective interaction mechanism [6,7]. However, the three tasks are interrelated from the perspective of semantic information. Knowing the viewpoint of the evaluation object is very important for judging its emotion. The evaluation viewpoint is usually adjacent to the evaluation object, and knowing one of the viewpoints is also helpful for another extraction task, and multiple evaluation objects have different emotional attitudes.

Recently, researchers have borrowed deep learning model algorithms and used multi-modal sentiment lexical extraction to extract attributes of text words and perform sequential classification, which can greatly improve the efficiency of manual feature extraction and save workload. Deep learning has the function of local feature extraction of text content and can store memory, which is the advantage of deep learning, so more and more text sentiment analysis tasks are solved with deep learning. Deep learning is suitable for word processing and semantic understanding because it has structural flexibility, and word embedding can be used in the underlying architecture of deep learning to avoid the problem of uneven text length. Therefore, the deep learning method based on neural networks has become the main research direction of ABSA sentiment analysis.

One part of the convolutional neural network that is particularly different from other neural network structures is the region of the artificial neuron structure, which responds accordingly throughout the area covered by the surrounding units. Convolutional neural networks have undertaken a lot of work on aspect-based sentiment analysis tasks. Yan W et al. used a convolutional neural network (CNN) to extract n-gram information of different granularities from each sentence, and then, the text contextual semantic information features were extracted by integrating these sentences sequentially through bi-directional gated recursive units (BiGRU) [8]. However, the attention mechanism is not incorporated in the CNN-BiGRU model, and the results show that the classification accuracy and recall are lower than those of CapsNet-BERT. Lv Y et al. introduced deep memory networks, bidirectional long- and short-term memory networks, and multiple attention mechanisms [9] to better capture the same or different emotional polarity of different aspects of textual contexts, and this method was only applied to the car review dataset, and the model reference range was narrow. Pang G et al. used a pre-trained BERT model to mine aspect-level auxiliary information from comment utterances to construct an aspectual feature location model (ALM-BERT) in order to learn the expressive features of aspectual words and the interaction information of the context of aspectual words [10]. However, existing aspect-level sentiment analysis methods focus on attention mechanisms and re-current neural networks. They lack emotional sensitivity to the location of aspectual words and tend to ignore long-term dependencies. Therefore, this model can hardly reflect people's fine-grained and integrated emotions adequately, which leads people to make wrong decisions. Sun C et al. developed the Matrix-Interactive Attention Network (MIAN) [11] to capture multiple interactive representations of target and contextual contexts, and to identify the sentiment polarity of the expressed opinion-specific target.

The above deep learning method model based on a convolutional neural network can achieve the purpose of aspect-level sentiment word feature extraction, but it is difficult to capture the complex semantic and syntactic information in the text, resulting in the loss of aspect-level sentiment utterance information. Especially in the process of data preprocessing, it is not sensitive to the sentiment word information of sentences, which increases the configuration scale of parameters.

In response to the above problems, a baseline model CapsNet-BERT combining capsule network and BERT is proposed, and the SemEval 14 Restaurant Review dataset is annotated. Experiments show that this method outperforms existing baseline methods. The contributions of this paper are as follows:

1. The SemEval 14 Restaurant Review dataset is divided into two versions, ATSA and ACSA. For ATSA, we extract aspect terms from sentences and map them with the corresponding sentiment polarity, and then remove sentences with the same opinion

in one aspect or multiple aspects. For ACSA, each sentence is mapped to an aspect category and corresponds to the sentiment of that aspect category.

2. A baseline model, CapsNet-BERT, is proposed. By optimizing the dynamic routing algorithm and sharing global parameters, the relationship between the local features of text sentences and the overall sentiment polarity is obtained, avoiding the degradation of aspect-level sentiment analysis to single sentence-level sentiment analysis.
3. The baseline model CapsNet-BERT connects the aspect embedding with each word of the sentence embedding and feeds it back to the bidirectional GRU through the residual connections to obtain a contextual representation, which more accurately predicts the sentiment polarity of related sentences.

The rest of the paper is structured as follows. In Section 2, we provide an overview and analysis of the relevant methods of the paper. In Section 3, the dataset for the experiments, the algorithmic model and the experimental design are introduced. In Section 4, the experimental data are analyzed and experimental conclusions are drawn. In Section 5, a summary of the thesis is presented.

2. Related Works

In recent years, multimodal sentiment analysis has become a hot topic of current research. It is found that, there are affective regions that evoke human sentiment in an image, which are usually manifested by corresponding words in people's comments. Therefore, Zhu T et al. introduce a cross-modal alignment module to capture region-word correspondence [12]. An adaptive cross-modal gating module is used to fuse multimodal features so that the features of images and texts can be more adequately connected for the purpose of reliably predicting contextual information. A recent trend in research is unsupervised aspect-based sentiment analysis methods for textual reviews [13]. People express their opinions, views on exoskeleton technology through the Twitter web platform. Researchers have used the Twitter platform textual reviews to evaluate the effectiveness of exoskeleton technology methods more effectively. In the unsupervised aspect-based sentiment analysis approach, Latent Dirichlet allocation, along with linguistic rules, is used for aspect extraction. SentiWordNet lexicon is used for sentiment scoring and classification. Researchers have also proposed other algorithms and models to capture the mutual representations between aspects and contextual background. Capsule networks and BERT models are both relevant to this paper.

2.1. Augmenting ABSA with Capsule Networks

Feature-based sentiment analysis relies on the relationship between pre-trained word embeddings and text sequence aspects for model construction, making the experimental models much more effective [7,14]. However, this approach relies heavily on the quality of the text content, as well as the construction of text word embeddings and task-specific architectures. A text with multiple comment targets may exhibit multiple sentiment tendencies. Traditional deep learning approaches based on neural networks usually use pooling or attention mechanisms to find specific sentiment word polarities, which can effectively separate these sentiment features [15–17]. However, this is not enough to aggregate these sentiment words with different polarities to the corresponding targets. To solve this problem, a capsule network is used to construct vectorized features, and the EM dynamic routing algorithm is used to replace the pooling operation. Dynamic routing can use clustering to separate or superimpose intersecting features [18,19], which can strengthen the relationship between the target words and the context.

Capsule network it is a set of neurons and the output of a neuron is an activity vector to represent a whole or a part of a whole [20]. The capsule network extracts the implicit features of the context and aspect words through sequential convolution, while introducing an interactive attention mechanism to reconstruct the vector representation of the context and aspect words. The effectiveness of the capsule network was verified on the SemEval 2014 dataset, but the problem of feature overlap in single-sentence text could not be solved

in the process of passing parameters between capsule layers. The CapsNet-BERT proposed in this paper solves the feature overlap problem in single-sentence text to some extent by introducing high-level capsule coefficients to adjust the contribution of each lower-level capsule to the higher-level capsule when dealing with the aspect-level sentiment classification task. In addition, the traditional neural network approach is scalar-scalar, while the capsule network approach is vector-vector, that is, their feature vectors (activity vectors) have some directionality. The feature vector paradigm (length) based on the text context sentiment word attributes represents its confidence level, and each different dimension represents a different feature pattern. The structure of the capsule network is shown in Figure 1.

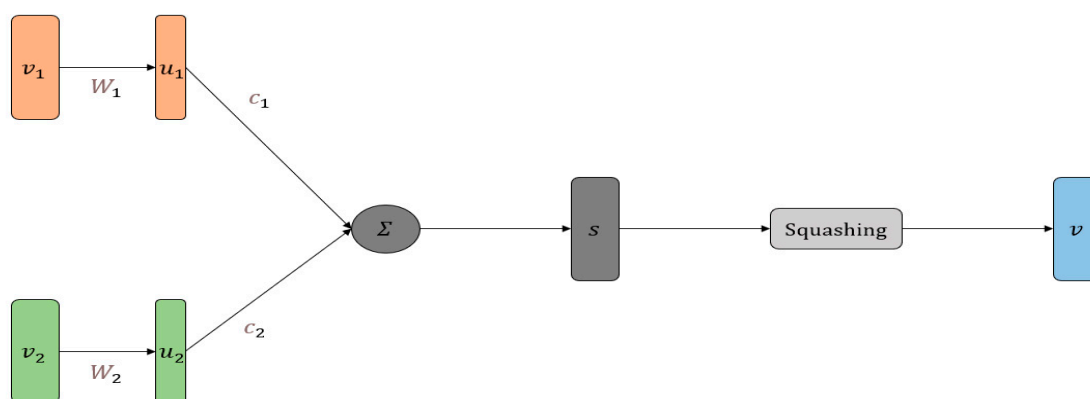


Figure 1. Representation of the structure of a capsule network.

The activity vectors v_1 and v_2 output from the upper layer of the capsule network are transformed into u_1 and u_2 by the affine transformation of W_1 and W_2 , and then S , is calculated according to Formula (1), and finally the output activity vectors are obtained by the squashing function Formula (2).

$$S = \sum_i u_i c_i \quad (1)$$

$$V = \frac{\|S\|^2}{1 + \|S\|^2} \frac{S}{\|S\|} \quad (2)$$

The capsule network uses the margin loss function to adjust the optimization parameters, calculated according to Formula (3), the time size is now taken as k , $T_k = 1$, otherwise $T_k = 0$. $m^+ = 0.9$, $m^- = 0.1$ is used to limit the value of the activity vector $\|v_k\|$, if the k class is true and the prediction result is false, then $T_k = 1$, $\|v_k\|$ will be smaller and L will be larger, and vice versa.

$$L_K = T_K \max(0, m^+ - \|v_k\|)^2 + \lambda (1 - T_k) \max(0, \|v_k\| - m^-)^2 \quad (3)$$

2.2. Using BERT Model to Study ABSA

Aspect-based sentiment analysis models do not perform well when dealing with relationships between contextual texts. BERT relies on a powerful bi-directional transformer encoder to centralize tasks downstream, then process them, and dynamically extract attributes of sentiment words in specific text sentences [21,22]. The traditional word vector tool (Word2vec) also only provides the word embedding function of text context words [23]. Unlike ELMo, which employs an LSTM cascade to produce downstream tasks [24], LSTMs need to be trained independently from left to right and right to left to get the task done [12,25].

BERT is a deep bidirectional unsupervised model of linguistic expressions and it only processes plain text corpus, where the relationship of text words can be contextual or

context-independent [24–26]. BERT uses the masked language model (MLM) and next sentence prediction (NSP) as unsupervised objects during pre-training, and the input representation is constructed from the corresponding word block embeddings containing more parameters, segment embeddings and positional embeddings are added together, so the vector representation of each word and phrase output by the model is more powerful and describes the overall information of the input text more comprehensively and accurately, providing better initial values for subsequent fine-tuning as much as possible.

Figure 2 shows the structure of the BERT model, in which the representation corresponding to the sentiment input words in each contextual text is constructed and then plugged into the bidirectional Transformer, which finally outputs a deep bidirectional linguistic representation that incorporates contextual information of equal length to the input. The advantages of the Transformer are summarized in at least three points.

- It breaks the limitation that RNN models cannot perform parallel operations on a large data corpus of textual contexts.
- In contrast to CNN, the number of operations required to compute the association between two locations does not increase with distance.
- In the attention mechanism, examine the content of the textual context word attention distribution. Each focus can learn to perform different tasks.

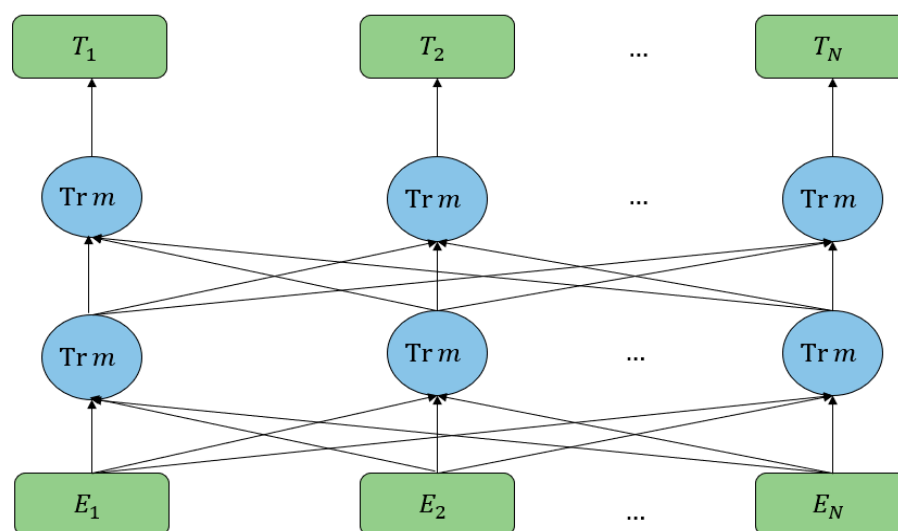


Figure 2. Bert pre-training language model.

3. Experimental Design

3.1. Datasets

The paper uses three evaluation datasets: two versions of ATSA and ACSA from the SemEval 14 Restaurant Review dataset, and the Laptop Review dataset. The SemEval 14 Restaurant Review dataset is labeled with two versions of ATSA and ACSA. It is characterized by a sentence which has at least two different aspects of sentiment polarity. It avoids the degradation of aspect-based sentiment analysis to sentence-level sentiment analysis. For ATSA, aspect terms in the sentences have been extracted and mapped to the corresponding sentiments, and any sentence with one or more aspects having the same sentiment is removed. the ACSA dataset is extracted for aspect sentiment categories.

The SemEval 14 Laptop Review dataset is only labeled with one version of ATSA, and the “Laptop” dataset has many implicit sample expressions in the context of the word text. However, the amount of specific data is very small, and it is very difficult to use.

In order to combine the capsule network and Bert model for correlation analysis study, the three sentiment corpus evaluation datasets were processed and divided into training and test sets. The tabular statistics of ATSA, ACSA, and Laptop review datasets are as follows: Tables 1–3.

Table 1. ATSA dataset table statistics.

	Attribute Word (Number)	Each Sentence Contains Attributes (Number)	Each Sentence Contains Aspects (Number)
Training set	3805	2475	2.61
Test set	955	619	2.62
Validation set	1898	1241	2.60

Table 2. ACSA dataset table statistics.

	Attribute Word (Number)	Each Sentence Contains Attributes (Number)	Each Sentence Contains Aspects (Number)
Training set	2459	1607	2.25
Test set	630	411	2.25
Validation set	1280	853	2.22

Table 3. Laptop review dataset table statistics.

	Attribute Word (Number)	Each Sentence Contains Attributes (Number)	Each Sentence Contains Aspects (Number)
Training set	3351	2134	2.52
Test set	1032	402	2.53
Validation set	1280	853	2.51

In the three evaluation datasets, the original data contains the labels of attribute words, aspect categories and sentiment polarity categories, where “Opinion target” and “Aspect term” both represent the meaning of sentiment analysis content attribute words. The difference is that each attribute word polarity or contextual sentiment word sentence aspect in the data has its own corresponding different aspect categories and different sentiment word attribute classification labels [27–29]. According to the different characteristics of contextual textual content lexicality, sentiment word polarity is divided into three judgment criteria: positive, negative, and neutral.

The goal of the experiment is to accurately identify the polarity of the sentiment category in the words of a comment sentence when the textual attribute words of the sentiment sentence are known.

3.2. Model Design

Given an aspect term or an aspect category of a sentence, we want to extract the lexical features of the target sentiment sentence, so the CapsNet-BERT model is proposed. CapsNet-BERT combines the size of the local feature representation generated by the sentiment lexicality of the text context with the feature representation related to the global aspect sentence, and optimizes the target sentiment sentences by a soft classifier to simulate learning the complex relationship between aspect and context.

The CapsNet-BERT model consists of four layers: embedding layer, encoding layer, primary capsule layer, and category capsule layer. It combines the advantages of BERT and a capsule network, we replace the embedding and encoding layers of CapsNet with pre-trained BERT, and the CapsNet-BERT model takes [CLS]sentence[SEP]aspect[SEP] as the input to get the embedding of each word directly. The model is shown in Figure 3.

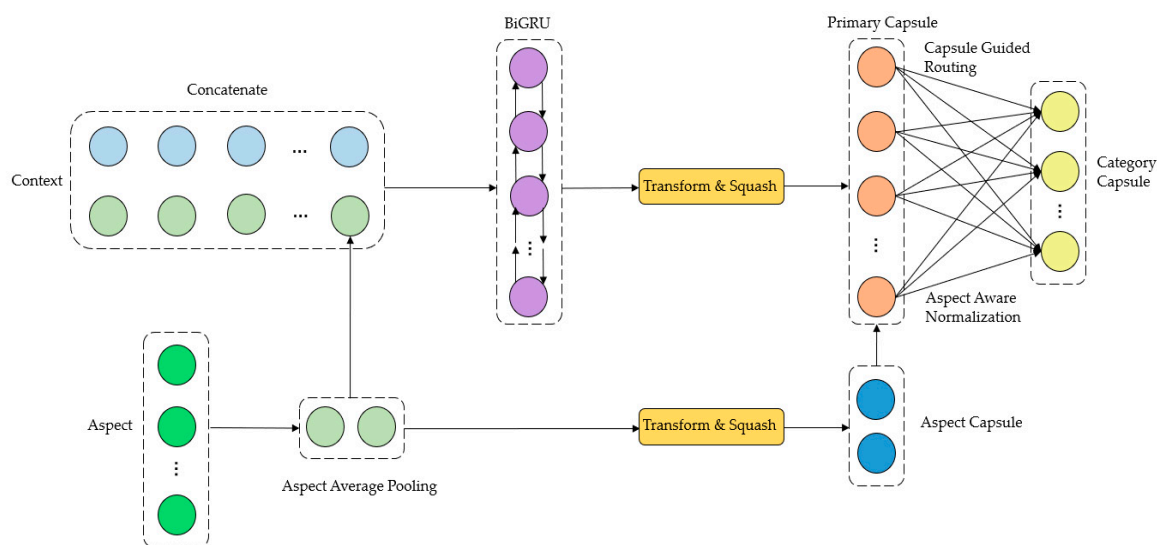


Figure 3. Represents the architecture of the CapsNet-BERT model.

3.3. Methodology

We denote the set of sentences in the training data by D . Given a sentence $S = \{w_1^s, \dots, w_n^s\}$, an aspect term $A^t = \{w_1^a, \dots, w_l^a\}$, an aspect category A^c , the aspect-level sentiment analysis aims to predict the sentiment polarity $g \in \{1, \dots, N\}$ of A^t or A^c in sentence S . Denote by w a specific word, l the length of the aspect term t , and N the number of sentiment categories.

3.3.1. Embedding Layer

At this layer, the input sentences and aspects are converted into word embeddings. When the task is ATSA, aspect embedding is performed according to the average value of aspect terms, which allows more semantic information to be stored intuitively [22,30]. When the task is ACSA, m is first initialized randomly and aspect embedding is followed by learning. The output of the embedding layer is a matrix embedding of sentiment word feature vector sentences with perceptual aspects, and we associate each word aspect embedding m in the sentence with S . As shown in Formula (4):

$$E_i^{sm} = [E_i; m] \quad (4)$$

3.3.2. Encoding Layer

The aspect category embedding obtained by learning may not fully capture the semantics, so try to embed words or introducing prior knowledge as a way to initialize the aspect category embedding [31]. Then, the aspect-aware word embedding (E^{sm}) is input into BiGRU to obtain the hidden state $H = [h_1, \dots, h_n]$ for each word corresponding to the contextual sentiment semantics. Then, using the role of the residual connection function in the model structure, the aspect-aware word embedding is added to all the structural hidden states, such as Formula (5):

$$H = \text{BiGRU}(E^{sm}) + E^{sm} \quad (5)$$

BiGRU and BERT mainly identify words related to the current aspect and embed aspect information into the representation of the words. The method of aspect embedding which carries the lexical features of aspect sentiment analysis is used so that the output of the trained BiGRU can capture the information of different words under the current aspect. Finally, the residual connection ensures that the aspect information is not lost in the model training.

3.3.3. Primary Capsule Layer

In the primary capsule layer, we obtain the primary capsule $P = [P_1, \dots, P_n]$ and the aspect capsule C by linear transformation and squeezing the activation function. Where W^p, b^p, W^a and b^a are the learning parameters. The calculation Formula is as (6), (7).

$$\text{primary capsule } P : P_i = \text{squash}(W^p h_i + b^p) \quad (6)$$

$$\text{aspect capsule } C : C = \text{squash}(W^a a + b^a) \quad (7)$$

The primary capsule layer is calculated and each BiGRU hidden state corresponds to produce a capsule, using the aspect embedding sentiment word feature attribute to produce an aspect capsule. here, the capsule is finally produced by a linear transformation, squeezing the activation function [32]. After making full use of the contextual representation method to get the main capsule P , the aspect ratio capsule is embedded using the aspect ratio from the embedding layer. There are two mechanisms in this layer:

- Aspect aware normalization, in order to overcome the problem of training instability due to variation in sentence length, we use aspect capsule to normalize the primary capsule weights [33]. When applied to text, the uncertainty of text length can lead to instability in the training process, which is a problem that needs to be addressed by the capsule network. Specifically, when the text is long, the number of primary capsules is large, which will cause the mold of the upper-layer capsule to become larger, resulting in a higher probability of each upper-layer superstructure capsule unit cell. Therefore, the standardization mechanism (scaling the values of a column of numerical features in the training set to a mean of 0 and a variance of 1) is used in the paper, and standardization in deep learning better maintains sample spacing. When there are outliers in the samples, normalization may “crowd” the normal samples together. For example, if there are three samples, and the value of a feature is 1, 2, 10,000, suppose 10,000 is an outlier, after using normalization, the normal 1 and 2 will be “squeezed” together. If the classification labels of 1 and 2 are opposite, then when we use gradient descent to train the classification model, it will take longer for the model to converge because it requires more effort to separate the samples, whereas standardization does a good job in this regard, at least it does not “squeeze” the samples together. Standardization is more in line with the statistical assumption that for a numerical feature, it is very likely to follow a normal distribution. Standardization is actually based on the implicit assumption that the normal distribution is adjusted to a standard normal distribution with mean 0 and variance 1. where W^t is the learning parameter, as shown in Formula (8):

$$u_i = \frac{\exp(p_i W^t c)}{\sum_{j=1}^t \exp(p_j W^t c)} \quad (8)$$

First each primary capsule in the capsule network is multiplied by the weight u , and then the average weight is calculated by the attention mechanism method [25], which just conceptually replaces the vector with a capsule. In addition, the attention mechanism here plays the effect of extracting feature vectors normalization by text sentiment analysis [16] and also plays the effect of soft attention by aspect to pick the primary capsule that is relevant and important to it and then to calculate the upper capsule. This is why this model works well.

- Capsule guided routing exploits the prior knowledge between sentiment contextual text attribute categories to improve the routing process [18,34]. In the training process, the sentiment matrix $G \in R^{c \times d}$ is calculated by averaging the manually collected word embeddings of similar sentiment words, and then the number of sentiment categories C is obtained. d is the dimensionality of the sentiment embedding. The input squeeze activation function to obtain the sentiment capsules $Z = [Z_1, \dots, Z_C]$,

and then the routing weight W is calculated by measuring the similarity between the primary capsule layer and the sentiment capsule.

As shown in Formulas (9) and (10).

$$Z_i = \text{squash}(G_i) \quad (9)$$

$$w_{ij} = \frac{\exp(p_i W^r z_j)}{\sum_{k=1}^n \exp(p_i W^r z_k)} \quad (10)$$

This method of routing is actually based on the existing results to find the corresponding primary capsule related to the polarity of words in the text context.

3.3.4. Category Capsule Layer

Using primary capsules, aspect-aware normalized weights and capsule-guided routing weights, the category capsule $U = [U_1, \dots, U_C]$, Equation (11) is finally calculated as.

$$\text{Category capsule } U_j = \text{squash}(q \sum_{i=1}^n w_{ij} u_i p_i) \quad (11)$$

where q is the learning ratio parameter for extending the connection weights.

We use margin loss as an aspect-based intelligent classification loss function [35], where $T_k = 1$ when and only when category k exists. m^+ and m^- are the boundary hyperparameters. In the experiments, m^+ , m^- and λ is set to 0.9, 0.1, and 0.6. Margin Loss is as in Equation (12):

$$L = \sum_{k=1}^C T_k \max(0, m^+ - \|v_k\|)^2 + \lambda (1 - T_k) \max(0, \|v_k\| - m^-)^2 \quad (12)$$

4. Experimental Details

4.1. Experimental Parameters

- (1) Based on aspect term (Aspect_term_model) the experimental parameters areas in Tables 4 and 5.

Table 4. Recurrent_capsnet setting parameters.

Recurrent_Capsnet	Values
embed_size	300
dropout	0.5
num_layers	2
capsule_size	300
bidirectional	True
optimizer	Adam
batch_size	64
learning_rate	0.0003
weight_decay	0
num_epochs	20

Table 5. Bert_capsnet setting parameters.

Bert_Capsnet	Values
bert_size	768
dropout	0.1
capsule_size	300
optimizer	Adam
batch_size	32
learning_rate	0.00002
weight_decay	0
num_epochs	5

- (2) Based on aspect category (Aspect_category_model) the experimental parameters are as in Tables 6 and 7.

Table 6. Recurrent_capsnet setting parameters.

Recurrent_Capsnet	Values
embed_size	300
dropout	0.5
num_layers	2
capsule_size	300
bidirectional	True
optimizer	Adam
batch_size	64
learning_rate	0.0003
weight_decay	0
num_epochs	20

Table 7. Bert_capsnet setting parameters.

Bert_Capsnet	Values
bert_size	768
dropout	0.1
capsule_size	300
optimizer	Adam
batch_size	32
learning_rate	0.00003
weight_decay	0
num_epochs	5

4.2. Experimental Evaluation Metrics

To evaluate the performance of the model for textual aspect-level sentiment analysis, we introduced accuracy and F1 measures as evaluation criteria.

4.2.1. Accuracy

The ratio of the number of samples correctly classified by the model to the total number of samples is the accuracy rate, and generally speaking, the higher the accuracy rate, the better. Formula (13) is described as follows (number of samples where *TN* true sentiment is negative and correctly predicted, number of samples where *FN* true sentiment is negative but incorrectly predicted, number of samples where *TP* true sentiment is positive and correctly predicted, and number of samples where *FP* true sentiment is positive but incorrectly predicted).

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (13)$$

4.2.2. Precision

The precision rate is the percentage of samples with a positive prediction result in which the true label is also positive. Formula (14) is as follows.

$$Precision = \frac{TP}{TP + FP} \quad (14)$$

4.2.3. Recall

Recall refers to the ratio of the number of samples in which the true sentiment is positive and the prediction is correct to the total number of samples. Equation (15) is described as follows.

$$Recall = \frac{TP}{TP + FN} \quad (15)$$

4.2.4. F1-Measure

Due to the mutual influence between the two evaluation index values of precision rate and recall rate, the optimal value cannot be achieved at the same time. In order to make a better overall evaluation of the model, the arithmetic mean of the precision rate and the recall rate is calculated to obtain the new evaluation index F_1 measure. The calculation Formula (16) is as follows.

$$F_1 = \frac{\text{Precision} * \text{Recall} * 2}{\text{Precision} + \text{Recall}} \quad (16)$$

4.3. Model Performance Experimental Comparison Analysis

The CapsNet-BERT model is compared with three types of models in other papers: LSTM-based is based on LSTM, CNN-based is based on CNN, and Attention-based is based on attention. Finally, the effectiveness of the CapsNet-BERT combination is compared by ablation studies. The paper evaluates the performance of the model with accuracy rate and F1 measures through both ATSA and ACSA versions in the SemEval 14 Restaurant Review dataset, and the Laptop Review dataset. The experimental results are shown in Table 8.

Table 8. Comparison of experimental results of aspect-level sentiment polarity analysis. %.

Models	ATSA		ACSA		Laptop Review	
	Accuracy	F1	Accuracy	F1	Accuracy	F1
TextCNN	51.62	51.53	48.79	48.61	75.85	75.23
LSTM	50.46	50.13	48.74	48.65	75.83	75.01
TD_LSTM	75.49	73.38	-	-	75.92	75.20
AT_LSTM	77.43	73.58	66.42	65.31	73.85	72.13
ATAE_LSTM	77.18	71.62	70.13	66.78	77.68	76.19
BiLSTM+Attn	76.18	70.62	66.25	66.16	77.47	77.12
AOA_LSTM	76.35	71.14	-	-	81.25	81.06
CapsNet	79.80	73.47	73.54	67.25	80.58	80.12
BERT	82.25	79.41	84.68	75.52	79.68	78.29
CapsNet-BERT	83.49	80.96	85.69	76.32	80.62	79.62
CapsNet-DR	79.89	71.36	69.18	65.32	76.78	76.12
CapsNet-BERT-DR	82.95	80.07	79.73	75.42	79.53	78.23

- As described in Table 8, the Laptop Review dataset outperformed ATSA and ACSA in the sentence sentiment classifiers TextCNN and LSTM, suggesting that ATSA and ACSA can avoid degrading the dataset into sentence-level sentiment analysis.
- The SOTA ABSA method on the Laptop Review dataset performs poorly on ATSA and ACSA, indicating that the ATSA and ACSA datasets are challenging to train the corpus for textual contextual affective lexicality.
- Attention-based models without proper modeling of textual word sequences do not perform well because the sequential information of sentences is lost, resulting in inadequate processing of the ATSA and ACSA datasets, and therefore fail to link context to aspects.
- On the datasets Laptop Review, ATSA, and ACSA, the combined CapsNet-BERT performs better than its text feature word extraction model.

4.4. Ablation Study

CapsNet-DR and CapsNet-BERT-DR are included only for measuring the effectiveness of capsule-guided wiring and are only used for simple comparison with CapsNet-BERT. If standardized dynamic routing (DR) is utilized, it is easy to reduce the performance of the model, and the performance of these two combinations is not as good as CapsNet-BERT.

5. Conclusions and Future Work

ABSA has become increasingly popular in recent years for a growing number of practical applications. For example, it can help improve consumer demand and guide

producers to improve their products. Aspect-level sentiment analysis is attempted to be applied in image sentiment classification. Specifically, on the one hand, we can roughly use the results of image saliency detection as the relative representation of different regions for sentiment classification and obtain the final feature representation by weighting the local features; on the other hand, we can let the network automatically determine the importance of different regions for sentiment classification by training the network as a whole, and then further obtain the weighted feature representation. These two methods, by optimizing the feature generation mechanism, enhance the feature representation, thus further improving the image sentiment classification accuracy.

Aspect-based sentiment analysis is divided into two subtasks, ACSA and ATSA. In this paper, we label the SemEval 14 Restaurant Review dataset into two versions, ATSA and ACSA, in which each sentence contains multiple aspects with different sentiment polarity, avoiding degrading aspect-level sentiment analysis to sentence-level sentiment analysis and promoting aspect-based sentiment analysis research. In addition, the CapsNet-BERT model, has achieved good results in sentiment analysis user review sentiment lexical feature extraction.

In the future, there are still some challenges in deeper semantic sentiment word attribute extraction, which have not been satisfactorily addressed in the current research. They include the problem of comparative level sentences, where it is difficult to judge whether the attribute of a certain aspect and the polarity of the comment attitude are preferred in comparative level textual contextual word feature extraction content related sentences. In conditional sentences, it is difficult to extract emotions from an unknown, unrealistic context, etc. In the future, we will continue research to develop models that can tap deeper hidden relationships between texts.

Author Contributions: J.T. and W.S. conceived the idea of this paper; J.T., M.X. and C.X. designed and completed the experiments; X.W. gave modifications to the experiments; J.T. wrote and revised the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the sub-project of the National Key R&D Program, Dark Web Intelligence Analysis and User Identification Technology (Grant No. 2017YFC0820702-3), and the National Language Commission key Project, Cross-Media Multilingual Public Opinion Information Processing Based on Big Data in Cyberspace (Grant No. ZDI135-96).

Acknowledgments: We thank the anonymous reviewers for their valuable feedback.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Madhoushi, Z.; Hamdan, A.R.; Zainudin, S. Aspect-based sentiment analysis methods in recent years. *Asia-Pac. J. Inf. Technol. Multimed.* **2019**, *7*, 79–96.
2. Moreno-Ortiz, A.; Salles-Bernal, S.; Orrequia-Barea, A. Design and validation of annotation schemas for aspect-based sentiment analysis in the tourism sector. *Inf. Technol. Tour.* **2019**, *21*, 535–557. [\[CrossRef\]](#)
3. Al-Smadi, M.; Al-Ayyoub, M.; Jararweh, Y.; Qawasmeh, O. Enhancing aspect-based sentiment analysis of Arabic hotels' reviews using morphological, syntactic and semantic features. *Inf. Process. Manag.* **2019**, *56*, 308–319. [\[CrossRef\]](#)
4. Tran, T.U.; Hoang, H.T.T.; Dang, P.H.; Riveill, M. Multitask Aspect-Based Sentiment Analysis with Integrated Bidirectional LSTM & CNN Model. *IAES Int. J. Artif. Intell.* **2020**, *9*, 1–7.
5. Lyu, C.; Chen, B.; Ren, Y.; Ji, D. Long short-term memory RNN for biomedical named entity recognition. *BMC Bioinform.* **2017**, *18*, 1–11. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Yadav, A.; Vishwakarma, D.K. Sentiment analysis using deep learning architectures: A review. *Artif. Intell. Rev.* **2020**, *53*, 4335–4385. [\[CrossRef\]](#)
7. Hemmatian, F.; Sohrabi, M.K. A survey on classification techniques for opinion mining and sentiment analysis. *Artif. Intell. Rev.* **2019**, *52*, 1495–1545. [\[CrossRef\]](#)
8. Yan, W.; Zhou, L.; Qian, Z.; Xiao, L.; Zhu, H. Sentiment Analysis of Student Texts Using the CNN-BiGRU-AT Model. *Sci. Program.* **2021**, *2021*, 8405623. [\[CrossRef\]](#)
9. Lv, Y.; Wei, F.; Cao, L.; Peng, S.; Niu, J.; Yu, S.; Wang, C. Aspect-level sentiment analysis using context and aspect memory network. *Neurocomputing* **2021**, *428*, 195–205. [\[CrossRef\]](#)

10. Pang, G.; Lu, K.; Zhu, X.; He, J.; Mo, Z.; Peng, Z.; Pu, B. Aspect-Level Sentiment Analysis Approach via BERT and Aspect Feature Location Model. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 5534615. [\[CrossRef\]](#)
11. Sun, C.; Lv, L.; Tian, G.; Liu, T. Deep Interactive Memory Network for Aspect-Level Sentiment Analysis. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **2020**, *20*, 1–12. [\[CrossRef\]](#)
12. Zhu, T.; Li, L.; Yang, J.; Zhao, S.; Liu, H.; Qian, J. Multimodal sentiment analysis with image-text interaction network. *IEEE Trans. Multimed.* **2022**. [\[CrossRef\]](#)
13. Pathik, N.; Shukla, P. Aspect Based Sentiment Analysis of Unlabeled Reviews Using Linguistic Rule Based LDA. *J. Cases Inf. Technol.* **2022**, *24*, 1–9. [\[CrossRef\]](#)
14. Naseem, U.; Razzak, I.; Musial, K.; Imran, M. Transformer based deep intelligent contextual embedding for twitter sentiment analysis. *Future Gener. Comput. Syst.* **2020**, *113*, 58–69. [\[CrossRef\]](#)
15. Shuang, K.; Yang, Q.; Loo, J.; Li, R.; Gu, M. Feature distillation network for aspect-based sentiment analysis. *Inf. Fusion* **2020**, *61*, 13–23. [\[CrossRef\]](#)
16. Li, W.; Qi, F.; Tang, M.; Yu, Z. Bidirectional LSTM with self-attention mechanism and multi-channel features for sentiment classification. *Neurocomputing* **2020**, *387*, 63–77. [\[CrossRef\]](#)
17. Zou, H.; Xiang, K. Sentiment classification method based on blending of emoticons and short texts. *Entropy* **2022**, *24*, 398. [\[CrossRef\]](#)
18. Camacho, D.M.; Collins, K.M.; Powers, R.K.; Costello, J.C.; Collins, J.J. Next-generation machine learning for biological networks. *Cell* **2018**, *73*, 1581–1592. [\[CrossRef\]](#)
19. Jagannath, J.; Polosky, N.; Jagannath, A.; Restuccia, F.; Melodia, T. Machine learning for wireless communications in the Internet of Things: A comprehensive survey. *Ad Hoc Netw.* **2019**, *93*, 101913. [\[CrossRef\]](#)
20. Zhang, W.; Tang, P.; Zhao, L. Remote sensing image scene classification using CNN-CapsNet. *Remote Sens.* **2019**, *11*, 494. [\[CrossRef\]](#)
21. Lian, Z.; Liu, B.; Tao, J. SMIN: Semi-supervised Multi-modal Interaction Network for Conversational Emotion Recognition. *IEEE Trans. Affect. Comput.* **2022**. [\[CrossRef\]](#)
22. Acheampong, F.A.; Nunoo-Mensah, H.; Chen, W. Transformer models for text-based emotion detection: A review of BERT-based approaches. *Artif. Intell. Rev.* **2021**, *54*, 5789–5829. [\[CrossRef\]](#)
23. Orkphol, K.; Yang, W. Word sense disambiguation using cosine similarity collaborates with Word2vec and WordNet. *Future Internet* **2019**, *11*, 114. [\[CrossRef\]](#)
24. Ji, C.; Wu, H. Cascade architecture with rhetoric long short-term memory for complex sentence sentiment analysis. *Neurocomputing* **2020**, *405*, 161–172. [\[CrossRef\]](#)
25. He, Z.; Wang, Z.; Wei, W.; Feng, S.; Mao, X.; Jiang, S. A Survey on Recent Advances in Sequence Labeling from Deep Learning Models. *arXiv* **2020**, arXiv:2011.06727.
26. Si, Y.; Wang, J.; Xu, H.; Roberts, K. Enhancing clinical concept extraction with contextual embeddings. *J. Am. Med. Inform. Assoc.* **2019**, *26*, 1297–1304. [\[CrossRef\]](#)
27. Vo, A.D.; Nguyen, Q.P.; Ock, C.Y. Semantic and syntactic analysis in learning representation based on a sentiment analysis model. *Appl. Intell.* **2020**, *50*, 663–680. [\[CrossRef\]](#)
28. Yue, L.; Chen, W.; Li, X.; Zuo, W.; Yin, M. A survey of sentiment analysis in social media. *Knowl. Inf. Syst.* **2019**, *60*, 617–663. [\[CrossRef\]](#)
29. Tembhurne, J.V.; Diwan, T. Sentiment analysis in textual, visual and multimodal inputs using recurrent neural networks. *Multimed. Tools Appl.* **2021**, *80*, 6871–6910. [\[CrossRef\]](#)
30. Kumar, A.; Narapareddy, V.T.; Srikanth, V.A.; Neti, L.B.M.; Malapati, A. Aspect-based sentiment classification using interactive gated convolutional network. *IEEE Access* **2020**, *8*, 22445–22453. [\[CrossRef\]](#)
31. He, R.; Lee, W.S.; Ng, H.T.; Dahlmeier, D. Exploiting Document Knowledge for Aspect-level Sentiment Classification. *Proc. 56th Annu. Meet. Assoc. Comput. Linguist.* **2018**, *2*, 579–585.
32. Deng, Y.; Lei, H.; Li, X.; Lin, Y.; Cheng, W.; Yang, S. Attention Capsule Network for Aspect-Level Sentiment Classification. *KSII Trans. Internet Inf. Syst.* **2021**, *15*, 1275–1292.
33. Wadawadagi, R.; Pagi, V. Sentiment analysis with deep neural networks: Comparative study and performance assessment. *Artif. Intell. Rev.* **2020**, *53*, 6155–6195. [\[CrossRef\]](#)
34. Sharma, T.; Kaur, K. Benchmarking Deep Learning Methods for Aspect Level Sentiment Classification. *Appl. Sci.* **2021**, *11*, 10542. [\[CrossRef\]](#)
35. Xu, H.; Liu, B.; Shu, L.; Yu, P.S. BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; Volume 1, pp. 2324–2335.