



Article

HAMCap: A Weak-Supervised Hybrid Attention-Based Capsule Neural Network for Fine-Grained Climate Change Debate Analysis

Kun Xiang * and Akihiro Fujii

Department of Science and Engineering, Hosei University, Tokyo 184-8584, Japan; fujii@hosei.ac.jp

* Correspondence: kun.xiang.2u@stu.hosei.ac.jp

Abstract: Climate change (CC) has become a central global topic within the multiple branches of social disciplines. Natural Language Processing (NLP) plays a superior role since it has achieved marvelous accomplishments in various application scenarios. However, CC debates are ambiguous and complicated to interpret even for humans, especially when it comes to the aspect-oriented fine-grained level. Furthermore, the lack of large-scale effective labeled datasets is always a plight encountered in NLP. In this work, we propose a novel weak-supervised Hybrid Attention Masking Capsule Neural Network (HAMCap) for fine-grained CC debate analysis. Specifically, we use vectors with allocated different weights instead of scalars, and a hybrid attention mechanism is designed in order to better capture and represent information. By randomly masking with a Partial Context Mask (PCM) mechanism, we can better construct the internal relationship between the aspects and entities and easily obtain a large-scale generated dataset. Considering the uniqueness of linguistics, we propose a Reinforcement Learning-based *Generator-Selector* mechanism to automatically update and select data that are beneficial to model training. Empirical results indicate that our proposed ensemble model outperforms baselines on downstream tasks with a maximum of 50.08% on accuracy and 49.48% on F1 scores. Finally, we draw interpretable conclusions about the climate change debate, which is a widespread global concern.

Keywords: natural language processing; big data; data augmentation; capsule neural network; reinforcement learning; attention mechanism



Citation: Xiang, K.; Fujii, A. HAMCap: A Weak-Supervised Hybrid Attention-Based Capsule Neural Network for Fine-Grained Climate Change Debate Analysis. *Big Data Cogn. Comput.* **2023**, *7*, 166. <https://doi.org/10.3390/bdcc7040166>

Academic Editor: Moulay A. Akhloufi

Received: 12 September 2023
Revised: 12 October 2023
Accepted: 13 October 2023
Published: 17 October 2023



Copyright: © 2023 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

Stakeholders gradually notice the potential to apply Natural Language Processing (NLP) tools to help understand and control climate change (CC) dynamics since large language models (LLM) have been deployed to different application scenarios with significant results. However, two classic issues remain worthy of discussion and under intense debate in NLP: (1) they are highly reliant on powerful hardware and difficult to be applied to edge devices; (2) obtention of large-scale high-quality annotated datasets is laborious. Moreover, when it comes to a specific domain like CC, the available domain training dataset is scarce.

Convolutional Neural Network (CNN) is well known for its remarkable results and the breadth of applications, but still has some drawbacks. For example, in NLP, it is mainly manifested in the inability to capture the long-distance dependencies in the sequence to the pooling operation. Hinton et al. first proposed Capsule Networks in 2011 [1] and specifically defined them to be a set of neurons with instance parameters which represent active vectors [2]. The proposal of Capsule Neural Networks brought significant merits for the traditional CNN.

However, in fine-grained sentiment classification, content often targets multiple entities and corresponding aspects. Opinion holders have various emotional expressions towards several aspects of the same entity, making CC debates full of complexity and contradictions. Aiming at this, we need to further mine the internal relationship between

the aspects and the context. On the other hand, in aspect-level sentiment analysis tasks, context far from the aspect may have adverse effects on the classification results. Hence, we must take special consideration of the partial context closer to the aspect while dealing with the global context.

Therefore, we propose a novel Hybrid Attention-based Capsule Neural Network. We utilize the improved hybrid attention mechanism to capture the internal semantic structure, and the feature relation between the aspect and the context. In addition, we define a local window size to pinpoint the aspect-related partial context region and process it with a Partial Context Mask (PCM) mechanism for modeling the strong association of the context.

Another merit of the proposed PCM mechanism is that it brings us an opportunity of obtaining huge amounts of homologous data. Linguistic words and phrases can often have multiple transformations, such as case sensitivities, word order, and Part of Speech (PoS). Therefore, we propose to randomly mask partial context with a PCM mechanism to obtain new data as a *Generator*, and a Reinforcement Learning (RL)-based *Selector* is designed to help automatically select and update high-quality data that are helpful for model training with the reward mechanism. Ablation experiment results also indicate the effectiveness of the PCM mechanism for improving the robustness and generalization ability of the model.

The contributions of our work are summarized as follows:

(1) This research employs Capsule Neural Networks to process vast intricate Climate Change (CC) debates from *r/subreddits*. In particular, we design a novel hybrid attention mechanism tailored for fine-grained level context. This approach is demonstrated to be effective on downstream tasks, which outperforms the baselines on State-of-The-Art (SoTA) results.

(2) To address the challenge of data scarcity in a specific domain and considering the characteristics of aspect-based sentiment analysis, we propose a novel data augmentation method which is a Reinforcement Learning-based *Generator–Selector* mechanism. Specifically, a partial context mask (PCM) mechanism is designed as a *Generator* to allocate weights and randomly mask the non-local away from the aspect sequence. The perturbations produced by this process are discriminated by a *Selector* which can dynamically select and guide high-quality augmented data from the *Generator* outputs.

(3) We draw strong interpretable conclusions about CC debates from different levels by analyzing the emotional expressions of users in five climate-related *r/subreddits*.

The remainder of this paper is organized as follows: Section 2 presents an approach overview of current works. Section 3 depicts the technical components of our approach. Section 4 carries out all the experiment results and implement details. Section 5 offers statistical explanations of topic mining. Finally, we conclude the paper with a summary and directions for future work in Section 6. Appendix C presents key notations and abbreviations mentioned in the paper.

2. Related Works

2.1. Capsule Neural Networks for NLP

Due to the unique nature of Capsule Neural Networks (hereinafter referred to as CapNets), they are currently more widely used and developed in CV, but the emergence of the CapNets has opened up new ideas for researchers in many fields. CapNets can effectively reduce overfitting and improve the accuracy of detection. For NLP tasks, CapNets have also been shown to handle the relationship between subject–verb–object more effectively than traditional CNNs. Zhao et al. [3] proposed a capsule compression and partial routing mechanism to improve the scalability of capsule networks. They validated their approach on multi-label text classification and question-answering downstream tasks. Hettiarachchi et al. [4] designed a novel model based on CapNets to detect type and target of offensive posts in social media and achieved excellent results on a SemEval-2019 task. Liu et al. [5] proposed a capsule network based on a transformer encoder model (CapTE) which was verified to be efficient for predicting stock movements. Du et al. [6] proposed to use capsule networks with a novel “routing-on hyperplane” dynamic routing mechanism to

construct the model, and hyperplanes are utilized to decompose each capsule to acquire the specific senses. Xiao et al. [7] proposed a multi-task learning architecture based on capsule networks with the advantages of capsules for feature clustering. The experimental results on six classification tasks indicate the effectiveness of the model and the algorithm helps to reduce the interference among tasks. Su et al. [8] combined XLNet and capsule network to address the challenge of aspect-based sentiment analysis (ABSA); a capsule network with a dynamic routing algorithm was utilized to extract the local and spatial hierarchical relations of the text sequence and yield its local feature representations. Lin et al. [9] first employed the conventional source–target attention to produce a timestep-specific source-side context vector and fed the vector into a novel dynamic context-guided capsule network (DCCN) for multimodal machine translation (MMT), which achieved superior results on Multi30K dataset. Verma et al. [10] enhanced the traditional Graph CNNs by applying a capsule idea to solve graph classification problems. Wu et al. [11] combined capsule vectors with a Siamese network to handle the global semantic features and better obtain the spatial position relationships of local features, which outperformed other baselines on six different public datasets. Goldani et al. [12] used capsule networks to challenge fake news detection and achieved significant results on ISOT and LIAR datasets. Chen et al. [13] first proposed a few-shot learning COVID-19 rumor detection model based on capsule networks (CNFRD), which effectively utilized capsule neural layers to summarize the historical data and then obtain a generalized class representation. By calculating the distance between samples, epidemic rumor samples were discriminated by the metric module. Experimental results indicated the model surpassed the baselines on rumor datasets and could effectively improve rumor detection performance.

CapNets have achieved considerable results on NLP tasks so far, especially in different level sentiment classification and relation extraction tasks [14–17]. However, currently, no work is being devoted to applying CapNets for CC debates analysis.

2.2. Reinforcement Learning-Based Data Augmentation

Data augmentation methods are widely invested in deep learning. Among these methodologies, generative data augmentation is one of the most exciting emerging ideas [18]. However, these kinds of augmentation methodologies always cooperate with other technologies since the generated data always intermingle and lack faithfulness. Reinforcement Learning (RL) is proven to be effective and has made remarkable advances in CV and NLP. RL involves using a reward function, which helps an agent learn to interact to maximize the reward and then guides the model decision making [19].

An exciting approach is Data Boost proposed by Liu et al. [20] which augments data through Reinforcement Learning-guided conditional generation and achieves significant performance on three diverse text classification tasks compared to six other prior text augmentation methods. Pan et al. [21] proposed transferring knowledge from some important discourse markers to augment the quality of the Natural Language Inference (NLU) model and then use RL to optimize a new objective function with a reward defined by the property of the NLI datasets to make full use of the label's information. Ye et al. [22] proposed a novel State-Augmented RL framework (SARL) for portfolio management to solve the data heterogeneity problem. Cao et al. [23] proposed a deep generative RL model, which addressed the challenge of an imbalance class by augmenting the dataset with hateful tweets and achieved considerable improvement of detecting hate speech. Chen et al. [24] proposed an end-to-end RL framework where the generator automatically generates massive and diverse antonymous sentences while the discriminator evaluates the quality of the sample and helps the generator iteratively generate higher-quality antonymous samples. Xiang et al. [25] proposed a climate change domain-adapted distilled language model which applied an RL-based data augmentation method. They randomly substituted words in the sentence with the same Part of Speech for generating a large-scale dataset and used a selector to help select and guide the generator to iterate automatically.

2.3. NLP Technologies for Climate Change

The debate around CC is intense and of global importance [26], and NLP is well-positioned to help study the dynamics of the large-scale and complex discourse on CC. However, within the NLP community, the amount of work conducted so far on CC remains limited.

Mallick et al. [27] proposed a weak supervision-based NLP approach that leverages semantic similarity between categories and documents of CC debates and also offered some relevant conclusions based on topic models. Schafer et al. [28] proposed a reflexive, integrative approach for computational research on CC communication. Schweizer et al. [29] combined NLP tools and citation networks for automated literature review for scientific assessment of CC scenarios. Luccioni et al. [30] applied NLP techniques to pinpoint the companies that divulge their climate risks and those that do not. Loureiro et al. [31] assessed the sentiments and emotions of the tweets related to CC in the U.K. and Spain with NLP tools and explored how these relate to different preferences and concerns about energy policies. Swarnakar et al. [32] thematically discussed how NLP techniques could be employed in CC research and exemplified four NLP methodologies to explore four different extension tasks.

3. Model

In this section, we depict details of components of our approach, following the workflow with the sequence of hybrid attention-based CapNets, Reinforcement Learning-based data augmentation, and model ensemble.

3.1. Hybrid Attention-Based CapNet

The emotions and opinions to be expressed in practical applications correspond to different entities and aspects, especially when dealing with ambiguous texts like CC-related debates.

For example, “The global impact of the greenhouse effect is undeniable, climate warming may make agriculture more feasible in otherwise extremely cold regions, while it is significantly threatening to residents and ecosystems in coastal urban areas”. In aspect-oriented fine-grained sentiment classification, content often targets multiple entities and contains multiple aspects. In the sentence, “greenhouse” is the entity, while it also contains two aspects of “agriculture” and “ecosystem”. Obviously, “feasible” expresses positive sentiment towards “agriculture”, while “threatening” expresses negative concern towards “ecosystem”.

Opinion holders have different emotional expressions towards different aspects of the same entity, making CC debates full of complexity and contradictions, and making it more difficult in aspect-oriented fine-grained sentiment analysis.

As Figure 1 shows, the partial context is obviously more important for the correct sentiment polarity classification with multiple aspects. For a specific aspect, words with far distance have a weak influence and may even bring noise interference, which negatively affects the correct identification of sentiment polarity.

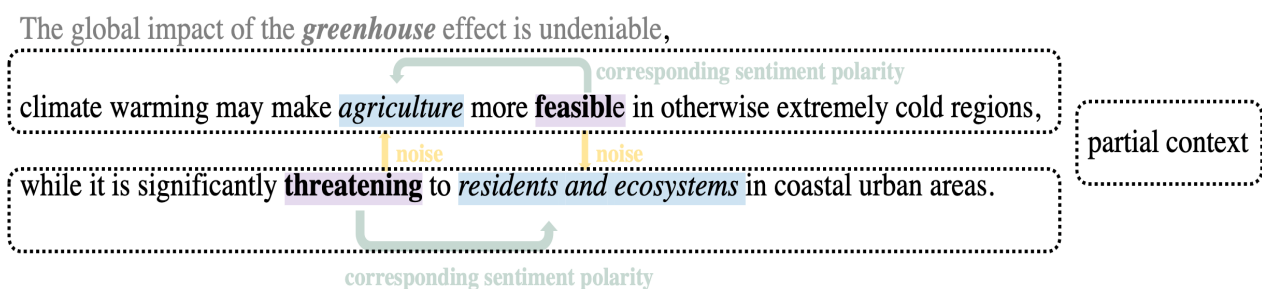


Figure 1. Example of content with multi-aspects in partial contexts.

In aspect-oriented fine-grained sentiment classification tasks, attention mechanisms are widely used to solve the problem of modeling specific attributes and their contextual relations. The attention mechanism can allocate additional weight information to the input features and hidden features for different sentiment aspects. However, aiming at CC-related texts, we need to further mine the internal relationship between the aspects and the context. On the other hand, in sentence-level sentiment analysis tasks, the traditional practice is to input all texts indiscriminately to obtain a comprehensive and thorough sentence representation. However, in aspect-level sentiment analysis tasks, context far from the aspect may have adverse effects on the classification results. Hence, we must take special consideration of the partial context closer to the aspect while dealing with the global context.

Therefore, we propose a Hybrid Attention Masking Capsule Neural Network (HAM-Cap). In essence, a Convolutional Neural Network is applied to construct the convolutional feature detector, extract n -gram information from the input word encoding sequence window and specify the input and output of the model and the intermediate data processing phase. Then, the improved hybrid attention mechanism is utilized to capture the internal semantic structure, as well as the feature relation between the aspect and the context.

In addition, we define a local window size to pinpoint the aspect-related partial context region and process it with a Partial Context Mask (PCM) mechanism for modeling the strong association of the context. Finally, we use the capsule network to classify the sentiment polarity and improve the routing algorithm and activation functions according to the specific tasks. The model structure is shown in Figure 2.

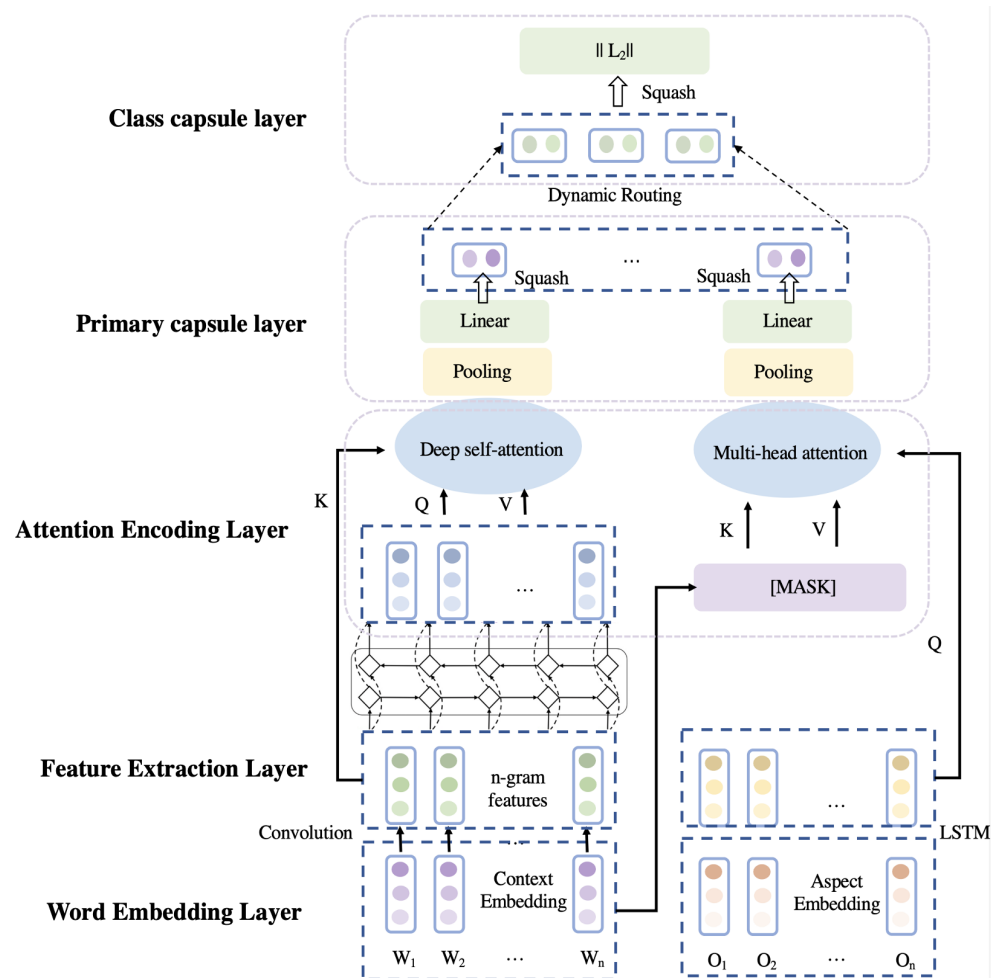


Figure 2. The network structure of the hybrid attention-masking capsule neural network (HAMcap).

3.1.1. Task Definition

Context sequence $s_c = \{w_1, w_2, \dots, w_n\}$ consisting of n words and an aspect sequence $s_a = \{o_1, o_2, \dots, o_i\}$ consisting of i aspects, where $o_k = \{w_k, w_{k+1}, \dots, w_{k+m-1}\}$ is a subsequence of s_c , are given. The target of aspect-based fine-grained sentiment analysis is to classify sentences based on different aspects, which can be expressed as

$$s_p = f_p(s_c, o_n), \quad (1)$$

where f is a nonlinear transformation function.

3.1.2. Word Embedding Layer

In the word embedding layer, a context sequence containing n words can be converted into $s_v = \{v_1, v_2, \dots, v_n\}$, where $v_i \in R^d$ is the d -dimensional vector representation of the i -th word, and $s \in R^{n \times d}$ represents the input word vector matrix of the sentence. Correspondingly, the aspect instance containing m words in the sentence is mapped to $T = \{v_n, v_{n+1}, \dots, v_{n+m-1}\}$, which is the aspect embedding sequence, where $v_i \in s$ is the d -dimensional vector representation of the i -th word of the aspect instance.

3.1.3. Feature Extraction Layer

The feature extraction layer applies multiple convolution operations to the input word vector matrix of the sentence to extract the corresponding n -gram features and generate a new feature vector matrix $M = m_1, m_2, \dots, m_{n-k+1}$, where $M \in R^{(n-k+1) \times d_k}$, k is the size of the one-dimensional convolution window, and d_k is the number of convolutional kernels.

Additionally, the LSTM network is applied to the aspect word embedding sequence to model the dependencies of words in attribute instances and mine their hidden semantics. Finally, the hidden state sequence $T_h = t_1, t_2, \dots, t_m$ obtained by the LSTM network is used as the high-level feature representation of the aspect word embedding sequence, where $T \in R^{m \times d}$, and q is the hidden layer dimensions of the LSTM network.

3.1.4. Hybrid Attention Mechanism

We accomplish attention encoding from two aspects. On the one hand, starting from the text itself, we focus on mining the internal structure of the context sequence; on the other hand, we investigate the aspect level, focusing on abstracting the semantic features of the context sequence and the specific aspect. The multi-head attention mechanism can simply and effectively abstract context dependencies and capture syntactic and semantic features.

This layer is deepened on the basis of standard multi-head attention. We design a deep self-attention mechanism and partial mask mechanism to further mine the input representation of the upper layer and generate two types of output features. Different from the traditional multi-head attention mechanism, this layer does not merge the sub-head results of the multi-head attention when outputting features, which is beneficial to expand the capsule types of primary capsules later.

Deep Self-attention Mechanism

In deep self-attention, the input feature sequence is first abstractly transformed, and the obtained high-level representation is added to the model to extend the standard self-attention mechanism. We utilize an LSTM network to perform further abstract operations on the input n -gram feature sequence G . The specific calculation process of deep self-attention is as follows:

$$DeepMHA = MH_{att}(G, H, H), \quad (2)$$

$$H = LSTM(G), \quad (3)$$

$$O^i = DeepMHA(G), \quad (4)$$

where $O^i \in R^{N \times (n-k+1) \times d}$ is the output of the deep self-attention.

Partial Context Mask Mechanism

In aspect-based sentiment analysis, the semantic relationship between the context sequence and the aspect instance is closely related to its relative position. However, traditional methods model the relationship between all input sequences and the aspect sequence indiscriminately. In order to emphasize the final impact of local context on sentiment polarity, a partial masking mechanism is designed to allocate weights according to the position of the context sequence and mask the non-local away from the aspect sequence.

In order to define the partial context of the aspect in the input sequence, we propose a partial context window (abbreviated to PCW and notated as P_{win}) to determine the boundary of the aspect, which is defined as follows:

$$P_{win} = |\lambda - L_\theta|, \quad (5)$$

$$L_\theta = \frac{1}{m} \sum_{i=\lambda}^{\theta+m-1} i, \quad (6)$$

where λ represents the location of the specific word v_λ of the partial context window, θ is the location of the first word in the aspect sequence, while m is the length of the aspect embedding sequence.

First, we construct the mask matrix $W^m = \{M_1, M_2, \dots, M_n\}$:

$$M_i = \begin{cases} E, & \text{if } |i - L_\theta| \leq P_{win} \\ 0, & \text{if } |i - L_\theta| \geq P_{win} \end{cases}, \quad (7)$$

where $E, 0 \in R_d$; then, we use the input context sequence S and the mask matrix W^m to perform element-wise operations on the corresponding position of the matrix to realize the partial context window mechanism and change the feature vectors falling outside the partial context window into the zero vector:

$$P_{win}(S) = S \otimes W^m. \quad (8)$$

This layer applies the partial context window to the input of the context sequence, generating weighted input feature sequence

$$V^m = P_{win}(S). \quad (9)$$

It is then combined with the upper layer input T_h to generate a multi-head attention high-level representation output

$$O^m = MHA(T_h, V^m), \quad (10)$$

where $O^m \in R^{N \times n \times d}$.

This layer also makes it possible to randomly generate a large amount of homologous data that paves the way for the subsequent data augmentation mechanism. We elaborate on details in the next chapter.

3.1.5. Capsule Neural Networks

Primary Capsule Layer

As the first layer of the capsule network, the primary capsule layer is responsible for encapsulating and processing two parts of the multi-head attention output O^i and O^m , converting it into a set of vector capsules that can be used by the class capsule layer.

Although the multi-head attention output is equipped with deep feature abstraction, it can only express local features and lacks sentence-level global semantic representation. This layer uses the global max pooling to compress the input of the upper layer in the

horizontal direction, so that the multi-head attention output features are aggregated in each subspace,

$$v_i^o = \text{globalmaxpooling}(o_i^c), \tag{11}$$

where $o_i^c \in o^g \cup o^m, v_i^o \in R^d$.

$$\text{compress} = \text{squash}(v_i^o W^c + b^c). \tag{12}$$

The squash function can compress the modulus length of the capsule vector to [0, 1], which is used to represent the existence probability of the feature. The definition is as follows:

$$\text{squash}(x) = \frac{\|x\|^2}{0.5 + \|x\|^2} \frac{x}{\|x\|}. \tag{13}$$

To enable the vector features in the capsule network to fully encode the sentence structure and semantic information contained in the context sequence, we use a variety of grain lexical combinations to expand the scale of the multi-head attention information subspace and enrich the semantic expression. Finally, this layer outputs the set of the primary capsule layer $P^c \in R^{4N \times d_c}$, where d_c is the dimensionality of the primary capsule:

$$P^c = \{p_1, p_2, \dots, p_{4 \times N}\}. \tag{14}$$

Class Capsule Layer

This layer is the final classification output layer composed of multiple types of capsules. Each type of capsule is a high-level representation of a classification, and the norm of each type of capsule indicates the probability that the original input sequence belongs to this category. The category with the highest probability is taken as the final classification result. In this work, for identifying the emotional polarity of the CC-related debate, we design three categories of capsules, corresponding to positive, negative, and neutral, respectively. The input and output of the capsule network take vector capsules as the basic unit, so it is different from the interaction between adjacent layers of the traditional DL networks. A dynamic routing is used between the layers of the capsule network to protocol the relationship. The upper-layer capsules iteratively aggregate toward high-level capsules until convergence.

To make the dynamic routing algorithm more effective, we introduce a weight transformation matrix between adjacent layers, which also makes the model have richer feature abstraction and combination capabilities.

We optimize and streamline the general dynamic routing algorithm specifically before starting the iteration. Capsule cap_i in the children capsule layer generates prediction vector V_{pre} for capsule cap_j in the parent capsule layer through the transformation matrix,

$$V_{pre} = cap_j W_i^c W \sim_j^c, \tag{15}$$

where $W_i^c \in R^{d_c \times d \sim c}$ is the weight transformation matrix corresponding to cap_j , $W \sim_j^c \in R^{d \sim c \times d \sim c}$ is the weight transformation matrix corresponding to V_{pre} .

Then, all the prediction vectors corresponding to the class capsule V_{pre} are weighted and summed to obtain the new vector representation of the class capsule, and enter the next iteration,

$$V_{pre} = \sum_i c_{ij} V \sim_{ij}, \tag{16}$$

where c_{ij} is a coupling function that indicates the aggregation strength of the bottom capsule to the high-level capsule. c_{ij} is obtained by applying the softmax function to the inner product of the prediction vector and the corresponding high-level capsule vector,

$$c_{ij} = \frac{\exp(b_{ji})}{\sum_k \exp(b_{ik})}. \tag{17}$$

When the iterative process is over, we bring u_j into the squash function to obtain the final output representation of capsule j , whose modulus length is limited to $[0, 1]$ to represent the activity probability of capsule j :

$$u_j^o = \text{squash}(u_j). \tag{18}$$

3.2. Reinforcement Learning Data Augmentation

This section introduces our proposed data augmentation approach which is a collaboration of *Generator* and Reinforced *Selector*. In the computer vision, there are affluent homologous images that are generated by synthesizing images through random clipping or random masking, which can not only augment data, but also improve the generalization ability of the model.

Since BERT [33] has been proposed, this kind of token mask and prediction mechanism is proven to be effective and even has been a default until now. Inspired by this, we propose a token replacement strategy which can also be called a partial context mask (PCM) mechanism-based *Generator* to automatically generate significant amounts of samples as Figure 3 shows. The text is usually a sequence in which each word is dependent on previous or following words; thus, it is crucial to learn the relationship weights among contexts. The partial context masking (PCM) mechanism is beneficial for modeling associations between words, semantic coherence of sentences and contextual information. The detailed formulations are demonstrated in Section 3.1.4.

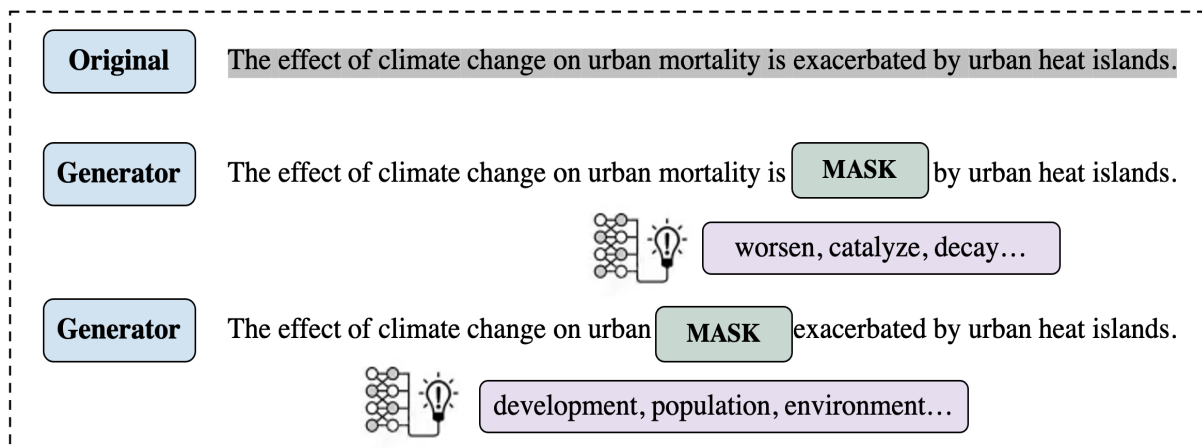


Figure 3. Random masking and prediction for token replacement-based Generator.

First, we are given a dataset $(D, Y) = \{(S_1, y_1), (S_2, y_2), \dots, (S_n, y_n)\}$ and a trained classification model $M: S \rightarrow Y$. We assume that for an input pair $(S = [t_1, t_2, \dots, t_m], y)$, we want to generate a new sample S' such that $M(S') \neq y$. Furthermore, we stipulate S' be similar to S in both grammar and semantics.

Another merit of the proposed PCM mechanism is that it brings us an opportunity of obtaining huge amounts of homologous data. Linguistic words and phrases can often have multiple transformations, such as case sensitivities, word order, and Part of Speech (PoS). This process brings perturbations to the sentences by masking a part of the tokens and using an LM to fill in the mask as the figure shows. As mentioned before, considering the characteristics of linguistics, in which a nuance may change the semantics of the sen-

tence, we tailor a *Reinforced Selector* based on Reinforcement Learning (RL) to dynamically select and guide high-quality augmented data from the *Generator* outputs. The selection criteria are based on assessing whether the chosen sample can improve the performance on validation and automatically update it according to the *reward*. The specific structure is demonstrated in Figure 4.

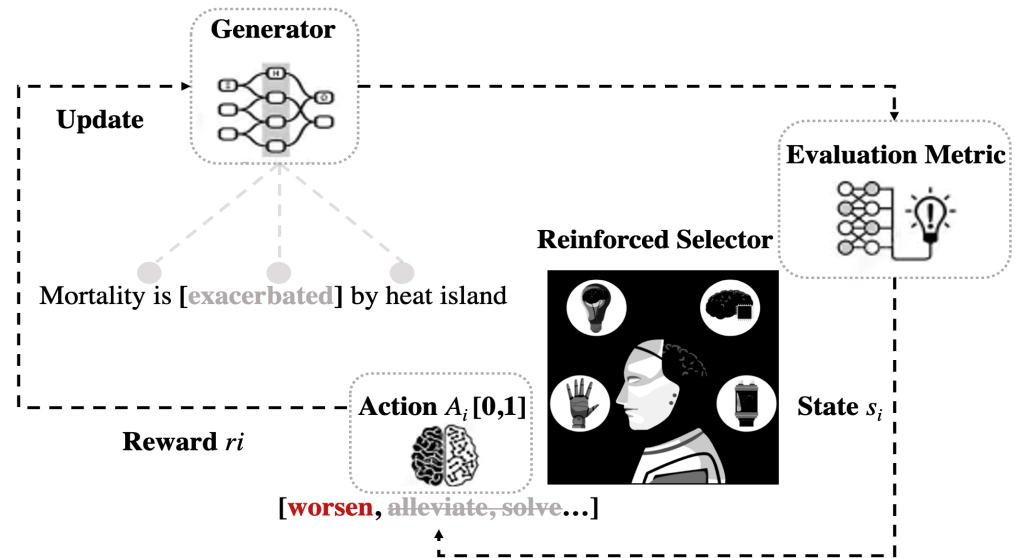


Figure 4. Reinforcement Learning-based selector mechanism.

4. Experiments

This section reports the dataset utilized for experiments, baselines, and implementation details.

4.1. Dataset

The dataset used in this work is based on a publicly available copy of Reddit (<https://www.reddit.com> accessed on 7 January 2023), all the posts and comments from all subreddits downloaded from Google Big Query data warehouse (<https://cloud.google.com/bigquery> accessed on 7 January 2023), spanning from January 2019 to October 2022. To identify terms related to CC topics, we follow the operating mode as in Mohammad et al. [34], after removing rare terms and those mainly used in other contexts on Reddit. In total, 11 terms are left for indexing CC-related discourses: "carbon dioxide", "carbon footprint", "carbon tax", "climate change", "climate model", "emission", "fossil fuel", "global warming", "greenhouse gas", "renewable", and "sea level". Among the 11 indexes, ignoring those smaller ones, we then identify 5 subreddits that focus on discussing CC as Table 1 shows: *r/climate*, *r/environment*, *r/climatechange*, *r/climateskeptics*, and *r/ClimateOffensive*.

The splitting ratio for each *r/subreddit* is 75% for training and 25% for testing. Specifically, we use the dataset consisting of 4375 hand-selected comments before augmenting for validation. All paragraphs are annotated as negative (risk), positive (opportunity), or neutral. The software used for collecting annotations is Prodigy (<https://prodi.gy> accessed on 19 April 2023). The detailed annotation rules are explained in Appendix A.

Table 1. Statistics of climate change-related posts and corresponding comments in 5 subreddits. We note that there are many empty posts and empty comments in one post, which are only with images or emojis, or some of them are deleted by themselves or moderators. Hence, we only count valid posts and comments, corresponding to Non-empty Posts and Comments, respectively.

Subreddits	<i>r/climatechange</i>	<i>r/climateskeptics</i>	<i>r/climate</i>	<i>r/ClimateOffensive</i>	<i>r/environment</i>
Non-empty Posts	1798	2121	2847	863	5899
Comments w/o DA	58,201	208,423	74,274	10,394	579,245
Comments w/ DA	79,327	381,645	93,071	17,394	839,572

4.2. Implement Details

4.2.1. Parameter Settings

All the experiments are carried out on a single Nvidia 16 GB V100 GPU. The dropout rate is 0.1 and batch size is either 32 or 64. The maximum sequence length is 100 and the L_2 regularization is $1e - 4$. The number of convolutional kernels is 250 and the hidden dimension is 300. The output capsule dimension is 16 and the iteration of dynamic routing is 7. The multi-head attention is set to be 8. For the RL policy network, the hidden layer is set to be 128. We use the Adam optimizer for Reinforced *Selector*, with $\beta_1 = 0.9$, $\beta_2 = 0.999$, respectively. Please refer to Appendix B for more details.

4.2.2. Baselines

(1) TransCap [35]. An aspect routing approach under transfer learning framework, the aim of which is to encapsulate the sentence-level semantic representations into semantic capsules from both aspect-level and document-level data.

(2) IAN [36]. An interactive attention network that inputs context information and aspect-embedding into two LSTM networks in order to obtain important information from an aspect-related attention mechanism.

(3) IARM [37]. A model that integrates relevant information from adjacent aspects for ABSA tasks. GRU and attention mechanism combined with a memory network to independently generate aspect-aware sentence representations for all aspects and repeatedly match the target aspect representation with other aspects to generate a more accurate target aspect representation.

(4) MemNet [38]. This model combines the attention mechanism with a deep memory network, steadily improving the classification accuracy of the model by superimposing multiple computing layers.

(5) RAM [39]. This algorithm optimizes the MemNet structure with a bi-LSTM network, combining recurrent neural networks and multiple attention mechanisms simultaneously in order to enhance the ability of obtaining long-dependency semantic features.

(6) ATAE-LSTM [40]. This model combines the attention mechanism and the LSTM network, which first splice aspect vectors and input features, then calculate the attention weights of the hidden state sequence, allocating different weights, and, finally, output.

(7) PRET + MULT [41]. Two approaches that transfer knowledge from document-level data, which is much more efficient and less expensive in order to improve the performance of aspect-level sentiment classification.

(8) BAT [42]. This is a novel network structure that combines adversarial learning and BERT pre-training.

(9) BERT-PT [43]. This model explores an improved post-training method based on a pre-trained BERT and fine-tunes it with a specific dataset for adapting downstream tasks.

4.3. Experimental Results

Obviously, our model consecutively outperformed most of the baselines on five *r/subreddits* as Table 2 shows. Its accuracy surpassed other baselines by 33.86% to 50.08%, and its F1 score surpassed other baselines by 33.31% to 49.48%. For the sake of fairness,

the baselines we selected all applied the attention mechanism. Not surprisingly, the ATAE-LSTM, which contains the shallowest model depth, performed the worst. This indicates that effectively increasing the network depth is important to improve model performance.

Table 2. Performance of sentiment analysis on five *r/subreddits*. The best results are highlighted with ♠, and the second best ones are marked with underlines.

Model	<i>r/climatechange</i>		<i>r/climateskeptics</i>		<i>r/climate</i>		<i>r/ClimateOffensive</i>		<i>r/environment</i>	
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
IAN	69.73	63.45	68.32	66.19	70.23	67.99	70.11	68.43	69.14	65.80
ATAE-LSTM	63.13	61.97	70.23	68.23	70.88	68.12	69.83	66.42	63.46	61.78
IARM	77.31	75.67	74.14	71.56	79.23	77.24	72.28	70.13	69.10	66.43
MemNet	79.13	77.23	80.13	74.32	82.45	79.13	80.53	78.13	81.34	77.63
RAM	82.13	80.44	83.13	80.34	80.13	76.34	83.14	80.42	79.13	77.48
TransCap	80.13	79.24	80.68	78.41	81.49	80.24	79.13	77.14	81.45	79.34
PRET+MULT	69.42	67.31	70.31	70.14	75.42	70.14	74.24	70.24	74.25	70.11
BAT	87.24	85.42	89.24	86.76	88.35	86.43	87.42	83.53	87.64	84.75
BERT-PT	89.35	87.53	89.96	♠87.50	90.64	89.53	91.54	♠89.64	91.54	89.34
HAMCap (ours)	♠92.42	♠89.43	♠91.45	<u>87.35</u>	♠94.23	♠90.64	<u>90.35</u>	<u>87.54</u>	♠95.24	♠92.35

Our model was designed based on the capsule network, so that the context can be abstracted at a higher level, while the ability to capture partial context features was optimized by the hybrid attention mechanism. It is worth noting that our model performed slightly worse on the two *r/subreddits*: *r/ClimateOffensive* and *r/climateskeptics*. This is mainly due to the fact that these two datasets have not yet reached a sufficient amount of data after data augmentation, which also demonstrates our Reinforcement Learning-based data augmentation strategy does play an important role in training models with complex data such as CC debates. This can also be seen from the fact that our model performed best on *r/environment*, because after data augmentation, this *r/subreddit* had the largest amount of data.

4.4. Analysis

4.4.1. Hybrid Attention Capsule Neural Network

Ablation Study

To investigate the effects of different components in our model, we conduct the following ablation study on our model. (i) “-Conv” denotes ablating convolution operations from the feature extraction layer, using the original input sequence features instead of n-gram features. (ii) “-HAM” denotes replacing the hybrid attention mechanism with a normal attention mechanism. (iii) “-PCM” denotes removing the partial context masking mechanism from the attention encoding layer, which makes the model ignore the partial context weight information of the input sequence. (iv) “-Cap” denotes substituting the capsule networks with a multilayer perceptron (MLP), the weighted features are directly passed to the softmax layer for classification.

We ablate the phases of critical technologies stage by stage. From Table 3, as expected, results for the simplified models all drop to a certain extent. This obviously demonstrates the necessity of all the components of our proposed model. Specifically, “-Cap” and “-PCM” affect the original model the most, since these are two essential mechanisms for building robust and precise connections between features and sentiment polarities among the context and the partial context.

Table 3. Ablation experiment of different components of our model. ↓ denotes the drop of performance. The worst scores are highlighted with ♣. We conduct the ablation experiments on three *r/subreddits*, which are the top three largest datasets after data augmentation. “Ori.” denotes the complete body of the proposed model.

<i>r/subreddit</i>	<i>r/climatechange</i>		<i>r/climate</i>		<i>r/environment</i>	
	Acc.	F1	Acc.	F1	Acc.	F1
Ori.	92.42	89.43	94.23	89.64	95.24	92.35
-Conv	4.34↓	6.45↓	5.25↓	7.52↓	5.76↓	♣8.49↓
-HAM	7.73↓	8.89↓	7.35↓	♣9.21↓	7.32↓	9.35↓
-PCM	8.77↓	9.45↓	♣9.54↓	8.35↓	6.34↓	8.56↓
-Cap	♣8.95↓	9.66↓	6.44↓	7.34↓	8.44↓	8.86↓

Partial Context Window Settings

To further verify the effectiveness of the partial context mechanism in improving the performance of the model, we investigate the influence of the size of the context window setting on the classification accuracy. A series of comparative experiments are carried out on three subreddits with different values of the window size to observe interpretable regularity. According to the principle of continuously expanding the window, the preset value is set to be [1, 10], which corresponds to the associated semantic area from small to large. To maintain quantitatively, the rest of the hyperparameters remain unchanged.

Figure 5 indicates that the optimal values of PCW are different depending on the datasets, and the classification accuracy is optimal at a specific partial context boundary, and there is a significant downward trend outside the boundary. Therefore, the design for partial context features can directly improve model performance, but it is necessary to select the most appropriate window size according to different datasets; otherwise, it brings disturbing noise by the obvious ambiguity and redundant semantic features, and leads to the risk of overfitting.

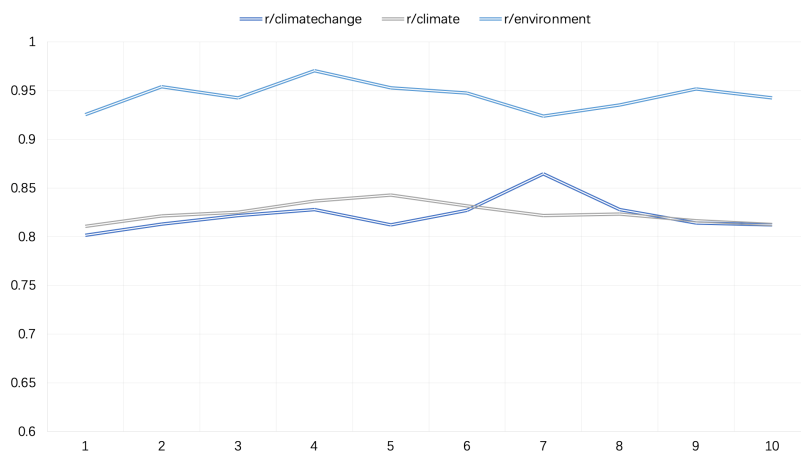


Figure 5. Classification performance on three subreddits with different PCW settings.

4.4.2. Data Augmentation Mechanism Analysis

Comparative Experiments on Effectiveness

To demonstrate the effectiveness of our proposed RL-based data augmentation, we conduct a series of comparative experiments of different DA algorithms. (i) “w/o sele.” denotes removing the RL-based selector mechanism, that is, only utilizing BERT’s masking and prediction mechanism to accomplish synonym substitution. (ii) “UDA” denotes unsupervised data augmentation proposed by Xie et al. [44]. “EDA” denotes easy data augmentation proposed by Wei et al. [45].

As shown in Table 4, our *Generator–Selector* collaboration data augmentation method outperforms other baselines in three subreddits. This indicates our data augmentation method is efficient in processing this kind of complicated dataset. Unsurprisingly, EDA performs the worst in this case. This accounts for easy data augmentation that only extends the dataset by adding noise directly, which makes the already ambiguous text more incomprehensible and difficult to capture the semantics. Unsupervised data augmentation utilizes a training signal annealing to alleviate overfitting of small-scale labeled datasets. However, the effect of UDA shows a declining trend with data increasing.

Table 4. Comparative experiment of the Reinforcement Learning-based data augmentation algorithm. “Collab.” denotes our proposed *Generator–Selector* collaboration data augmentation mechanism. “↓” denotes the performance loss.

<i>r/subreddit</i>	<i>r/climatechange</i>		<i>r/climate</i>		<i>r/environment</i>	
	Acc.	F1	Acc.	F1	Acc.	F1
Collab.	93.28	90.11	92.57	88.96	94.66	92.37
<i>w/o sele.</i>	7.53↓	6.45↓	7.13↓	5.73↓	9.45↓	7.48↓
UDA	2.56↓	4.89↓	4.35↓	2.21↓	5.32↓	3.35↓
EDA	14.73↓	16.25↓	12.51↓	13.79↓	15.40↓	16.56↓

Furthermore, it is obvious that by only using a masking and prediction mechanism to accomplish word substitution without the collaboration of the RL-based *Selector*, the model performance also drops to a certain extent. This accounts for the RL-based *Selector* being essential to help select and automatically upgrade the *Generator*.

Impact of RL Hyper-parameters

We demonstrate the impact of two hyper-parameters on our approach: the size of the hidden layer of the policy network and the reward discount factor. The numbers of units in the hidden layer of the policy network are 32, 64, 128, 256, 512. The choices for the reward discount factor are 0, 0.2, 0.4, 0.6, 0.8, 1.

Figure 6 presents the performance with different hyper-parameters. The performance of the model achieves the peak when the hidden layer is turned to 128 units. This indicates the importance of the capacity of the policy network which benefits the data selection. Furthermore, we notice from the figure that the reward factors do not affect the model performance too much. We interpret it as the fact that the reinforced selector can derive greater merits when it places proper emphasis on previous actions by applying relatively substantial reward discount factors but not related in particular.

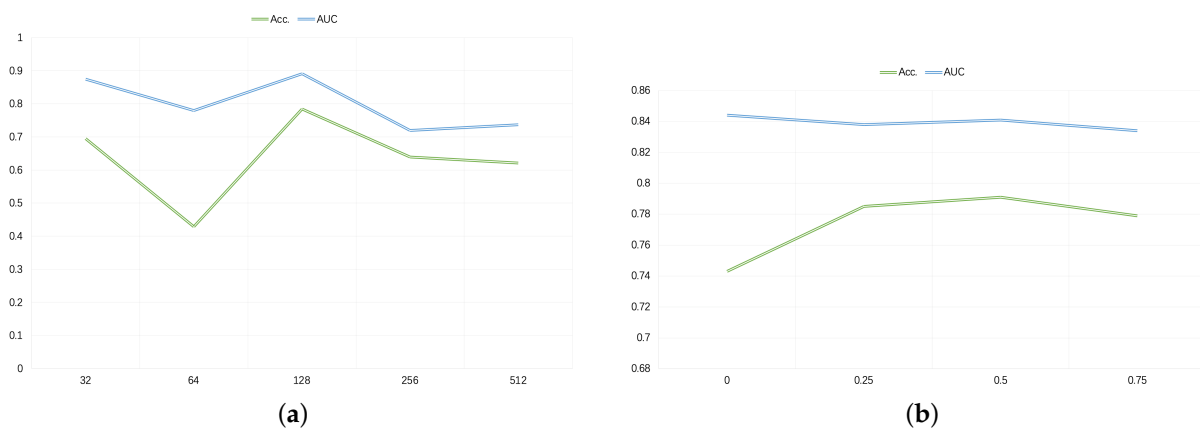


Figure 6. Impact of hyper-parameters of the Reinforced *Selector*. (a) Size of the hidden layer of the policy size; (b) Reward discount factor.

Figure 8 demonstrates different topic mining results based on positive and negative sentiment polarities. Nevertheless, our findings did not reveal a clear distinction in topics between the two sentiment polarities. In other words, the perspectives of both groups did not diverge significantly. This is a noteworthy result, suggesting that CC-related debates are always being presented in the form of conflict. This can be interpreted by the fact that stakeholders face a situation where risks and opportunities coexist to a certain extent.



Figure 8. Topic mining results of different sentiment polarities.

6. Conclusions

In this work, we proposed a novel hybrid attention masking model based on Capsule Neural Networks called HAMCap and Reinforcement Learning to address the data scarcity problem in NLP. To be more precise, we defined a local window size to pinpoint the aspect-related partial context region and process it with a partial context mask mechanism for modeling the strong association of the context.

Furthermore, to address the challenge of data scarcity, we proposed a novel data augmentation method which combines a partial context mask mechanism *Generator* and a reinforcement learning-based *Selector*. The experimental results fully demonstrated the effectiveness of our approach.

Although our approach achieved significant results on SoTA, due to the nature of CapNets, the inference speed of the model increased exponentially. With the increasing size of LMs, the consumption of training neural networks also increased exponentially. This is definitely a plight that we should consider particularly. In the future, we may consider combining HAMCap and knowledge distillation (KD) technology. Ideally, the model should break the limitations of edge devices and without losing model performance. In addition, this research applied CC-related datasets, and when we generalized it to other domains, the results declined to a certain extent.

Author Contributions: Conceptualization, K.X.; Methodology, K.X.; Software, K.X. and A.F.; Validation, K.X.; Formal analysis, K.X.; Investigation, K.X. and A.F.; Resources, K.X.; Data curation, K.X.; Writing—original draft, K.X.; Writing—review and editing, K.X.; Visualization, K.X.; Supervision, A.F.; Project administration, A.F.; Funding acquisition, A.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are not publicly available due to confidentiality.

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

Appendix A

Sampling comments and corresponding sentiment polarity labels.

Post: "I'm afraid climate change is going to kill me! Help!"		
Positive	"Since the end of the last ice age sea levels rose by 120 m and temperatures warmed by a minimum of 4 degrees but maybe as much as 7. We not only survived but civilization as we know it emerged. And that was before we had advanced technology. We will be ok".	Climate change is an irresistible result of natural development, whether it is caused by the development of modern society or not. Climate change is not going to harm or kill humans, and technological developments will hinder this dire trend.
Negative	"This post is 3 years old and still being used as an argument that 'everything is fine'. Way to disillusion yourself".	Believing that climate change is destroying the environment and human beings and showing a continuous deterioration.
Positive	"The climate of the earth is in constant change no matter what the cause be it volcano, radiation from space, solar fluctuations etc. There are thousands of things that will kill you that are greater than climate change such as car accidents, burning candles at home, smoking, drugs etc. Climate change danger from CO ₂ is much less than most household dangers so you need to get this all in perspective. Sure we must do things to prevent climate change the best we can but over population of the earth is a much bigger threat to man's existence. Too much population for us to produce food for!"	Admitting that climate change is ongoing and has some adverse effects. This change is brought about by the development of human society and will continue to pose threats. Therefore, measures need to be taken for intervention and protection. This deterioration is controllable.
Neutral	"Scientists said that the youngers are more delayed effect in comparison with the older people for sensitivity on NO ₂ and AMI".	Only states scientific data and expresses no personal perspective.

In the downstream task, we employed manually chosen paragraphs for both training and validation. Our annotation process adhered to the following guidelines: Annotators were tasked with categorizing paragraphs as either positive or negative alternatively, and also just as making a neutral statement.

Specifically, we provided a clear definition for categorizing paragraphs. A paragraph was labeled as “negative” if it depicted that climate change is occurring due to human activities, societal advancements, or modern industrialization, resulting in adverse effects on public health, serious consequences, or a worsening trend. In cases where a paragraph pertained to economic and business matters, the concept of “greenwash” was associated with a negative sentiment.

On the other hand, a paragraph was considered to express “positive” sentiment if it discussed climate change as a natural developmental phenomenon unrelated to human progress or societal activities, asserting that it will not impact human livelihoods. Additionally, if a paragraph highlighted climate change as offering potential opportunities to certain stakeholders or emphasized entities taking positive actions to address the issue, it was also classified as conveying positive sentiment.

Lastly, a paragraph fell under the “neutral” category if it objectively stated factual information or statistics without taking any particular perspective or stance of any stakeholder into account.

Appendix B

Specific experimental parameter settings.

Parameter	Value
Dropout	0.1
Batch_size	32/64
Maximum-sequence-length	100
Learning rate	1e-3/2e-5
L2-regularization	1e-4
Convolutional kernel numbers	250
Hidden layer dimension	300
Output capsule dimension	16
Dynamic routing iteration	7
Multi-head attention	8
Optimizer	Adam
Adam_beta1	0.9
Adam_beta2	0.999

Appendix C

Notation and abbreviation list.

Notation and Acronym	Explanation
w/	With
w/o	Without
r/	Specific reddit or subreddit
CC	Climate change
CapNets	Capsule neural networks
DA	Data augmentation
CV	Computer Vision
PCM	Partial context mask mechanism
RL	Reinforcement Learning
LLM	Large language model

References

1. Hinton, G.E.; Krizhevsky, A.; Wang, S.D. Transforming auto-encoders. In *Artificial Neural Networks and Machine Learning—ICANN 2011*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 44–51.
2. Sabour, S.; Frosst, N.; Hinton, G.E. Dynamic routing between capsules. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2017; Volume 30.
3. Zhao, W.; Peng, H.; Eger, S.; Cambria, E.; Yang, M. Towards scalable and reliable capsule networks for challenging nlp applications. *arXiv* **2019**, arXiv:1906.02829.
4. Ranasinghe, T.; Hettiarachchi, H. *Emoji Powered Capsule Network to Detect Type and Target of Offensive Posts in Social Media*; INCOMA Ltd.: Varna, Bulgaria, 2019.
5. Liu, J.; Lin, H.; Liu, X.; Xu, B.; Ren, Y.; Diao, Y.; Yang, L. Transformer-based capsule network for stock movement prediction. In *Proceedings of the First Workshop on Financial Technology and Natural Language Processing*, Macao, China, 12 August 2019; pp. 66–73.
6. Du, C.; Sun, H.; Wang, J.; Qi, Q.; Liao, J.; Wang, C.; Ma, B. Investigating capsule network and semantic feature on hyperplanes for text classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China, 3–7 November 2019; pp. 456–465.
7. Xiao, L.; Zhang, H.; Chen, W.; Wang, Y.; Jin, Y. Mcapsnet: Capsule network for text with multi-task learning. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, 31 October–4 November 2018; pp. 4565–4574.
8. Su, J.; Yu, S.; Luo, D. Enhancing aspect-based sentiment analysis with capsule network. *IEEE Access* **2020**, *8*, 100551–100561. [[CrossRef](#)]
9. Lin, H.; Meng, F.; Su, J.; Yin, Y.; Yang, Z.; Ge, Y.; Zhou, J.; Luo, J. Dynamic context-guided capsule network for multimodal machine translation. In *Proceedings of the 28th ACM International Conference on Multimedia*, Seattle, WA, USA, 12–16 October 2020; pp. 1320–1329.
10. Verma, S.; Zhang, Z.-L. Graph capsule convolutional neural networks. *arXiv* **2018**, arXiv:1805.08090.
11. Wu, Y.; Li, J.; Wu, J.; Chang, J. Siamese capsule networks with global and local features for text classification. *Neurocomputing* **2020**, *390*, 88–98. [[CrossRef](#)]
12. Goldani, M.H.; Momtazi, S.; Safabakhsh, R. Detecting fake news with capsule neural networks. *Appl. Soft Comput.* **2021**, *101*, 106991. [[CrossRef](#)]
13. Chen, D.; Chen, X.; Lu, P.; Wang, X.; Lan, X. Cnfrd: A few-shot rumor detection framework via capsule network for COVID-19. *Int. J. Intell. Syst.* **2023**, *2023*, 2467539. [[CrossRef](#)]
14. Du, C.; Sun, H.; Wang, J.; Qi, Q.; Liao, J.; Xu, T.; Liu, M. Capsule network with interactive attention for aspect-level sentiment classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China, 3–7 November 2019; pp. 5489–5498.
15. Yang, M.; Zhao, W.; Chen, L.; Qu, Q.; Zhao, Z.; Shen, Y. Investigating the transferring capability of capsule networks for text classification. *Neural Netw.* **2019**, *118*, 247–261. [[CrossRef](#)]
16. Fei, H.; Ji, D.; Zhang, Y.; Ren, Y. Topic-enhanced capsule network for multi-label emotion classification. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2020**, *28*, 1839–1848. [[CrossRef](#)]
17. Deng, J.; Cheng, L.; Wang, Z. Self-attention-based bigru and capsule network for named entity recognition (2020). *arXiv* **2002**, arXiv:2002.00735.
18. Shorten, C.; Khoshgoftaar, T.M.; Furht, B. Text data augmentation for deep learning. *J. Big Data* **2021**, *8*, 1–34. [[CrossRef](#)]
19. Jiang, J.; Chen, X.; Huang, Z.; Li, X.; Du, Y. Deep reinforcement learning-based approach for rumor influence minimization in social networks. *Appl. Intell.* **2023**, *53*, 20293–20310. [[CrossRef](#)]
20. Liu, R.; Xu, G.; Jia, C.; Ma, W.; Wang, L.; Vosoughi, S. Data boost: Text data augmentation through reinforcement learning guided conditional generation. *arXiv* **2020**, arXiv:2012.02952.
21. Pan, B.; Yang, Y.; Zhao, Z.; Zhuang, Y.; Cai, D.; He, X. Discourse marker augmented network with reinforcement learning for natural language inference. *arXiv* **2019**, arXiv:1907.09692.
22. Ye, Y.; Pei, H.; Wang, B.; Chen, P.-Y.; Zhu, Y.; Xiao, J.; Li, B. Reinforcement-learning based portfolio management with augmented asset movement prediction states. In *Proceedings of the AAAI Conference on Artificial Intelligence*, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 1112–1119.
23. Cao, R.; Lee, R.K.-W. Hategan: Adversarial generative-based data augmentation for hate speech detection. In *Proceedings of the 28th International Conference on Computational Linguistics*, Barcelona, Spain, 8–13 December 2020; pp. 6327–6338.
24. Chen, H.; Xia, R.; Yu, J. Reinforced counterfactual data augmentation for dual sentiment classification. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Online, 7–11 November 2021; pp. 269–278.
25. Xiang, K.; Fujii, A. Dare: Distill and reinforce ensemble neural networks for climate-domain processing. *Entropy* **2023**, *25*, 643. [[CrossRef](#)] [[PubMed](#)]
26. Stede, M.; Patz, R. The climate change debate and natural language processing. In *Proceedings of the 1st Workshop on NLP for Positive Impact*, Bangkok, Thailand, 5 August 2021; pp. 8–18.

27. Mallick, T.; Bergerson, J.D.; Verner, D.R.; Hutchison, J.K.; Levy, L.-A.; Balaprakash, P. Analyzing the impact of climate change on critical infrastructure from the scientific literature: A weakly supervised nlp approach. *arXiv* **2023**, arXiv:2302.01887.
28. Schäfer, M.S.; Hase, V. Computational methods for the analysis of climate change communication: Towards an integrative and reflexive approach. *Wiley Interdiscip. Rev. Clim. Chang.* **2023**, *14*, e806. [[CrossRef](#)]
29. Schweizer, V.J.; Kurniawan, J.H.; Power, A. Semi-automated literature review for scientific assessment of socioeconomic climate change scenarios. In Proceedings of the Companion Proceedings of the Web Conference 2022, Virtual Event, Lyon, France, 25–29 April 2022; pp. 789–799.
30. Luccioni, A.; Palacios, H. Using natural language processing to analyze financial climate disclosures. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019.
31. Loureiro, M.L.; Alló, M. Sensing climate change and energy issues: Sentiment and emotion analysis with social media in the UK and Spain. *Energy Policy* **2020**, *143*, 111490. [[CrossRef](#)]
32. Swarnakar, P.; Modi, A. Nlp for climate policy: Creating a knowledge platform for holistic and effective climate action. *arXiv* **2021**, arXiv:2105.05621.
33. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
34. Parsa, M.S.; Shi, H.; Xu, Y.; Yim, A.; Yin, Y.; Golab, L. Analyzing climate change discussions on reddit. In Proceedings of the 2022 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 14–16 December 2022.
35. Chen, Z.; Qian, T. Transfer capsule network for aspect level sentiment classification. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 547–556.
36. Ma, D.; Li, S.; Zhang, X.; Wang, H. Interactive attention networks for aspect-level sentiment classification. *arXiv* **2017**, arXiv:1709.00893.
37. Majumder, N.; Poria, S.; Gelbukh, A.; Akhtar, M.S.; Cambria, E.; Ekbal, A. Iarm: Inter-aspect relation modeling with memory networks in aspect-based sentiment analysis. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 3402–3411.
38. Tang, D.; Qin, B.; Liu, T. Aspect level sentiment classification with deep memory network. *arXiv* **2016**, arXiv:1605.08900.
39. Chen, P.; Sun, Z.; Bing, L.; Yang, W. Recurrent attention network on memory for aspect sentiment analysis. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 7–11 September 2017; pp. 452–461.
40. Wang, Y.; Huang, M.; Zhu, X.; Zhao, L. Attention-based lstm for aspect-level sentiment classification. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; pp. 606–615.
41. He, R.; Lee, W.S.; Ng, H.T.; Dahlmeier, D. Exploiting document knowledge for aspect-level sentiment classification. *arXiv* **2018**, arXiv:1806.04346.
42. Karimi, A.; Rossi, L.; Prati, A. Adversarial training for aspect-based sentiment analysis with bert. In Proceedings of the 2020 25th International conference on pattern recognition (ICPR), Milan, Italy, 10–15 January 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 8797–8803.
43. Xu, H.; Liu, B.; Shu, L.; Yu, P.S. Bert post-training for review reading comprehension and aspect-based sentiment analysis. *arXiv* **2019**, arXiv:1904.02232.
44. Xie, Q.; Dai, Z.; Hovy, E.; Luong, T.; Le, Q. Unsupervised data augmentation for consistency training. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 6256–6268.
45. Wei, J.; Zou, K. Eda: Easy data augmentation techniques for boosting performance on text classification tasks. *arXiv* **2019**, arXiv:1901.11196.
46. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.

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