2023 IEEE International Conference on Data Mining Workshops (ICDMW)

Knowledge Graph enhanced Aspect-Based Sentiment Analysis Incorporating External Knowledge

Autumn TEO[1], Zhaoxia WANG[1*], Haibo PEN[2*], Budhitama SUBAGDJA[1], Seng-Beng HO[3], Boon Kiat QUEK[3*]
School of Computing and Information Systems, Singapore Management University, Singapore, Singapore[1]
School of Electrical and Information Engineering, Tianjin University, Tianjin, China[2]
Social and Cognitive Computing Department, Institute of High Performance Computing (IHPC),
Agency for Science, Technology and Research (A*STAR), Singapore, Singapore[3]
autumn.teo.2018@scis.smu.edu.sg, {zxwang, budhitamas}@smu.edu.sg, penhaibo@tju.edu.cn, {hosb, quekbk}@ihpc.a-star.edu.sg

***Abstract*—Aspect-Based Sentiment Analysis (ABSA) is a type of sentiment analysis that could identify and extract various aspects or features from text and determining the sentiment associated with each aspect. ABSA has significant real-world applications, such as providing deeper insights about specific strengths and weaknesses of each aspect contained within text data. Despite notable advancements, ABSA still has room for improvement in completeness, accuracy, and performance efficacy. To tackle these challenges, this research introduces an approach to ABSA that leverages knowledge graphs to improve completeness, accuracy, and performance efficacy. Our key novelty is in being able to incorporate enhancements across multiple stages, including utilising knowledge graphs together with dataset processing, and architectural modelling. Additionally, we offer a complementary overview and analysis of various deep learning heuristics and optimization strategies that could further enhance ABSA performance. Our validation results demonstrate the effectiveness of the proposed knowledge graph enhanced ABSA method across multiple benchmark datasets, with notable boosts to model performance. Importantly, in being model-agnostic, our dataset processing approach could potentially enhance the performance of other ABSA methods in the future.**

***Index Terms*—Sentiment Analysis, Knowledge Graph, Aspect Based Sentiment Analysis, Deep Learning, External Knowledge**

I. Introduction

Sentiment analysis aims to classify people's opinions, sentiments, emotions and attitudes towards certain entities such as services, products or topics [1], [2]. Aspect Based (or level) Sentiment Analysis (ABSA) takes this to a finer-grained level, where the sentiment polarity of every target entity (or aspect) present is determined. Here, *polarity* refers to the *valence* of the sentiment, and this could either be *positive*, *neutral*, or *negative*, while *aspect* would refer to the particular attribute, category, feature, or topic that is being referenced [3], [4]. For example, in the sentence "*the screen is very clear, but the battery life is too short*", the entity "*screen*" has a positive sentiment, whereas "*battery life*" is associated with a negative sentiment.

*Corresponding Author

Recent approaches to ABSA, such as those incorporating Deep Neural Networks (DNN), have started to take *context* into account [5]. Among DNNs, Long Short-Term Memory (LSTM) networks have shown efficacy in uncovering the semantic relationship(s) between the target and its surrounding contextual information [6]. LSTMs could be further improved by incorporating attentional mechanisms to better capture important contextual information pertaining to a specific aspect [6]. To augment ABSA models further, recent attempts that incorporated external knowledge have started to outperform current state-of-the-art models [7]–[9]. However, these attempts either require substantial workarounds to integrate external knowledge, thus sacrificing generalizability, or the Knowledge Graph (KG) may be augmented inefficiently [3]. For example, in Zhao and Yu's approach [8], when the KG is queried to generate relevant triples for each input, the same triple could be produced multiple times, as the same topic could be present throughout the dataset, leading to inefficiency in KG augmentation. There are other approaches that utilize Bi-LSTM, which increases the complexity and training time.

In this paper, we present an LSTM-based model that could achieve greater accuracy but without sacrificing efficiency in the incorporation of KGs and external knowledge. Additionally, we demonstrate how performance could be further enhanced through various optimization techniques. Key contributions of this research includes:

- An enhanced ABSA method that leverages a KG and external knowledge. This method incorporates enhancements across various stages, encompassing the utilization of KGs, dataset processing, and architectural modelling, ultimately contributing to the advancement of ABSA.
- A comprehensive overview of various deep learning heuristics and optmization strategies, such as Optimizer, Learning Rate Scheduler, Weight Initializer, and Non-Linear Activation Function, which we employ, to push the overall ABSA task performance.

2375-9259/23/$31.00 ©2023 IEEE
DOI 10.1109/ICDMW60847.2023.00107

- A demonstration of the model's effectiveness, through evaluation on the Restaurant14 and Laptop14 datasets. Our model performs competitively against recent state-of-the-art models, which underscores the practical efficacy and advancements that could be made possible by the proposed approach.

II. Related Work

A. Sentiment Analysis

Sentiment analysis, also known as opinion mining, is a burgeoning field within natural language processing (NLP) that focuses on deciphering and understanding human sentiments or emotions expressed in textual data [4], [10]. In an era dominated by digital communication, sentiment analysis plays a pivotal role in gauging the subjective tone, attitude, and feelings embedded in vast amounts of textual content. Its applications span across diverse domains, ranging from social media platforms, where users openly share their opinions, to customer reviews of products and services, and even in the analysis of news articles [11]–[13]. The primary goal of sentiment analysis is to automatically determine whether a piece of text expresses a positive, negative, or neutral sentiment, offering valuable insights for businesses, researchers, and policymakers seeking to comprehend public opinion and sentiment trends [14].

In recent years, sentiment analysis has become increasingly popular for social data analysis [15]. Different AI techniques have been leveraged to improve both accuracy and interpretability of sentiment analysis algorithms, including symbolic AI [16], [17], subsymbolic AI [18], [19], and neurosymbolic AI [20], [21]. Besides traditional algorithms that focus on English text [22], multilingual [23] and multimodal sentiment analysis [24] have recently attracted increasing attention. Typical applications of sentiment analysis include social network analysis [25], [26], finance [27], and healthcare [28].

The methodologies employed in sentiment analysis have evolved over time, from early rule-based approaches to more sophisticated machine learning and deep learning models [29]. Traditional methods often relied on predefined rules or sentiment lexicons, while contemporary techniques harness the power of neural networks to capture intricate patterns and context in language [14], [30]. As sentiment analysis continues to grow in importance, researchers are exploring innovative ways to address challenges such as multilingual sentiment analysis [23], and aspect-based sentiment analysis [4], [5], [31], [32].

Aspect-based sentiment analysis (ABSA) is an advanced form of sentiment analysis that goes beyond the traditional classification of text into positive, negative, or neutral sentiments [4], [32]. Instead, ABSA aims to identify and analyze the sentiment associated with specific aspects or components within a piece of text. This fine-grained approach allows for a more nuanced understanding of opinions, especially in scenarios where multiple aspects or features are discussed [4], [5], [31], [32]. Various DNN-based methodologies have been employed to tackle ABSA tasks.

B. Aspect Based Sentiment Analysis

DNN-based models have been the focus of recent ABSA research with most models being either Convolutional Neural Networks (CNNs)-based [33]–[36], LSTM-based [37]–[41] models or a hybrid of both [42], [43]. The main motivating factor for using CNNs is its ability to extract local information from the data and synthesize the relationships among the features [44]. Another advantage of CNN is its non-linearity, enabling it to fit the data better than linear models and it also avoids the need for custom fixed language rules [34]. Toh and Su [35] achieved state of the art results using CNN on SemEval-2016 by combining a deep CNN model with a RNN model. Xu et al. [36] achieved a competitive result on the Yelp datasets by incorporating CNN with a non-linear Conditional Random Field (CRF) model to extract the aspect before using another CNN to predict the sentiment. However, CNN-based models often neglect important sequential information as they typically average the values of the aspect embeddings to obtain aspect information. Xu et al. [36] overcame this issue by using a CNN based module to improve target-specific representation. Another issue with CNN is the increase in complexity in proportion to the size of the dataset and the increase in the number of convolution layers can lead to a vanishing gradient problem [45]. Lastly, CNNs are also highly reliant on the initial parameters in order to avoid a local minima, requiring a considerable amount of work in initialization according to the problem at hand [46].

For textual data, it is advisable to utilize Recurrent Neural Networks (RNNs) instead. However, RNNs have a fatal flaw in which they are susceptible to the vanishing or exploding gradient problem where the backpropagated gradient tends towards zero and infinity respectively. For this reason, LSTMs are often used instead as it is able to mitigate this issue through its input, forget and output gates [46], [47]. Furthermore, LSTM-based models are able to use sequential information to capture long range semantic dependencies. For example, Li et al. [40] proposed LSTM model utilizes the positional dependencies of sentiment and aspect words. Tang et al's Target-Dependent LSTM (TD-LSTM) and Target-Connection LSTM (TC-LSTM) models extended upon the LSTM model by considering the target as a feature [48]. Wang et al.'s Attention-based LSTM (ATAE-LSTM) extends on TD-LSTM by utilizing an attention mechanism to utilize the relationships between aspects and polarity [49]. Ma et al. [39] proposed Interactive Attention Network (IAN) applied two attention networks to identify key words of the target and its full context. Chen et al. [38] introduced a recurrent attention network that employs a recurrent attention structure to capture aspect-specific sentence representation. Sun et al. [37] proposed a solution employing a dependency tree, Bi-LSTM and Graph Convolutional Network (GCN) to propagate contextual and dependency information between various aspects and opinions. Li et al. [41] used two LSTMs to capture history-aware-aspect representation before using an attention mechanism to transform it into sequential-aspect representations.

C. Knowledge Graph Providing External Knowledge

A KG contains information about various real-world entities, providing a rich source of external information for the model learning process [50]. Despite this, much of the existing research predominantly focuses on mining context-word-to-aspect-word dependencies within the sentence itself, largely neglecting the integration of text-related external knowledge, such as data on related words and symbolic knowledge [51]. The utilization of external knowledge has proven beneficial in enhancing semantic representation, leading to improved performance in sentiment analysis systems.

A KG is often denoted as a collection of triples (h, r, t), consisting of a head entity, $h \in E$, a tail entity, $t \in E$, and a relation, $r \in R$. Here E denotes $\{e_1, e_2, \ldots, e_n\}$ and it is a set of entities, R denotes $\{r_1, r_2, \ldots, r_n\}$ which represents a set of binary relations. An example of a triple is "(AlfredHitchcock, DirectorOf, Psycho)" where AlfredHitchock is the head entity and Psycho is the tail entity respectively while DirectorOf represents the relation [52]. DBpedia [53], YAGO [54], and Freebase [55] are examples of important KGs that query, store and represent relational real-world semantic data. Recent research has incorporated KGs to tackle a wide variety of tasks which includes information extraction [56], question answering [57] and semantic parsing [58].

However, the inherent nature of the triples poses a challenge when it comes to manipulating it for use in downstream tasks. This has given rise to KG embeddings as a means of addressing this issue. KG embeddings preserve its structure while simplifying the manipulation by embedding the entities and its relations within a vector space [52]. There are multiple KG embedding techniques which can be categorized into two different groups: semantic matching and translational distance models. As the name suggests, semantic matching models perform semantic matching of entities and relations to their vector space representation. RESCAL assigns every entity to a vector to capture its semantic meaning while representing each relation as a matrix before using a bilinear function to determine the plausibility of a fact [59].

ANALOGY extends on RESCAL by optimizing the analogical properties of the representations of the embedded relations and entities [60]. DistMult simplifies RESCAL [61], which decreases its usability as it is only able to handle symmetrical relationships, making it unsuitable for general KGs. Holographic Embedding (HolE) [62] combines the strength of RESCAL and DistMult through the use of circular correlation operation. Similarly, ComplEx extends on DistMult by utilizing complex-valued embeddings, enabling it to better model asymmetric relations [63].

The translational distance models utilize distance-based function to determine the plausibility of a fact. An example of a translational distance is TransE [64], which represents relations as translations in the embedding space. For a given triple (h, r, t), the h and t entities will be connected by a vector r, i.e., $h + r \approx t$.

However, TransE is only effective in dealing with 1-to-1 relations. TransH [65] overcomes this deficiency by allowing entities who are involved in multiple relations to have distinct representations. TransR [66] and TransD [67] further expand on the translational distance model by adopting relation-specific spaces rather than hyperplanes. However, TransM [68], TransF [69] and TransA [70] employ different methodologies as they relax the translational requirement $(h + r \approx t)$.

III. Knowledge Graph enhanced ABSA

This section presents the proposed method: KG enhanced aspect-based sentiment analysis. As shown in Fig. 1, the proposed model, inspired by KGAN [9], contains three sections: Context, Knowledge and Syntax sections to encode semantic features from multiple viewpoints. The knowledge section introduces external knowledge into the semantic features through the use of an external KG. The contextual and syntactic sections extract features from the word embedding representation of the dataset to establish the relevance relationship between the entity and the sentiment. Lastly, the output of each section is combined together to provide a complete depiction.

A. Context Section

The research done by Ding et al. [71] and Yang et al. [72] suggests that modelling contextual information can improve performance. Thus, we can employ pretrained word embedding models to numerically represent the vocabulary of the data. In our model, we utilized GloVe [73] and Bidirectional Encoder Representations from Transformers (BERT) [32] to perform word embedding and the initialization of the embedding matrix.

Word embedding embeds each word, w_i, within a vector space $E \in R^{(|V| \times d_w)}$ where $|V|$ refers to the dataset vocabulary and d_w refers to the word embedding dimension. Similar to Tang, Qin and Liu [74], we further encoded the relative position feature to take advantage of the positional information of aspect, entity and context words. As shown in Fig. 1, we leveraged two LSTMs to learn the relationship between the aspect and entity in each individual sentence. Additionally, two attention mechanisms were used to identify aspect-specific contextual information. The first attention mechanism learns any long-range dependencies of the contextual information while the second attention mechanism learns the weight of each aspect before aggregating the different weights to obtain the contextual representation of each aspect R_c.

B. Syntax Section

To leverage syntactic information, we utilized the same LSTM and pretrained word embedding that was used in the context section to obtain the hidden state vectors. This reduces the model complexity and training time. We used the hidden state vectors as input for a 2-layer GCN to capture the important syntactic information.

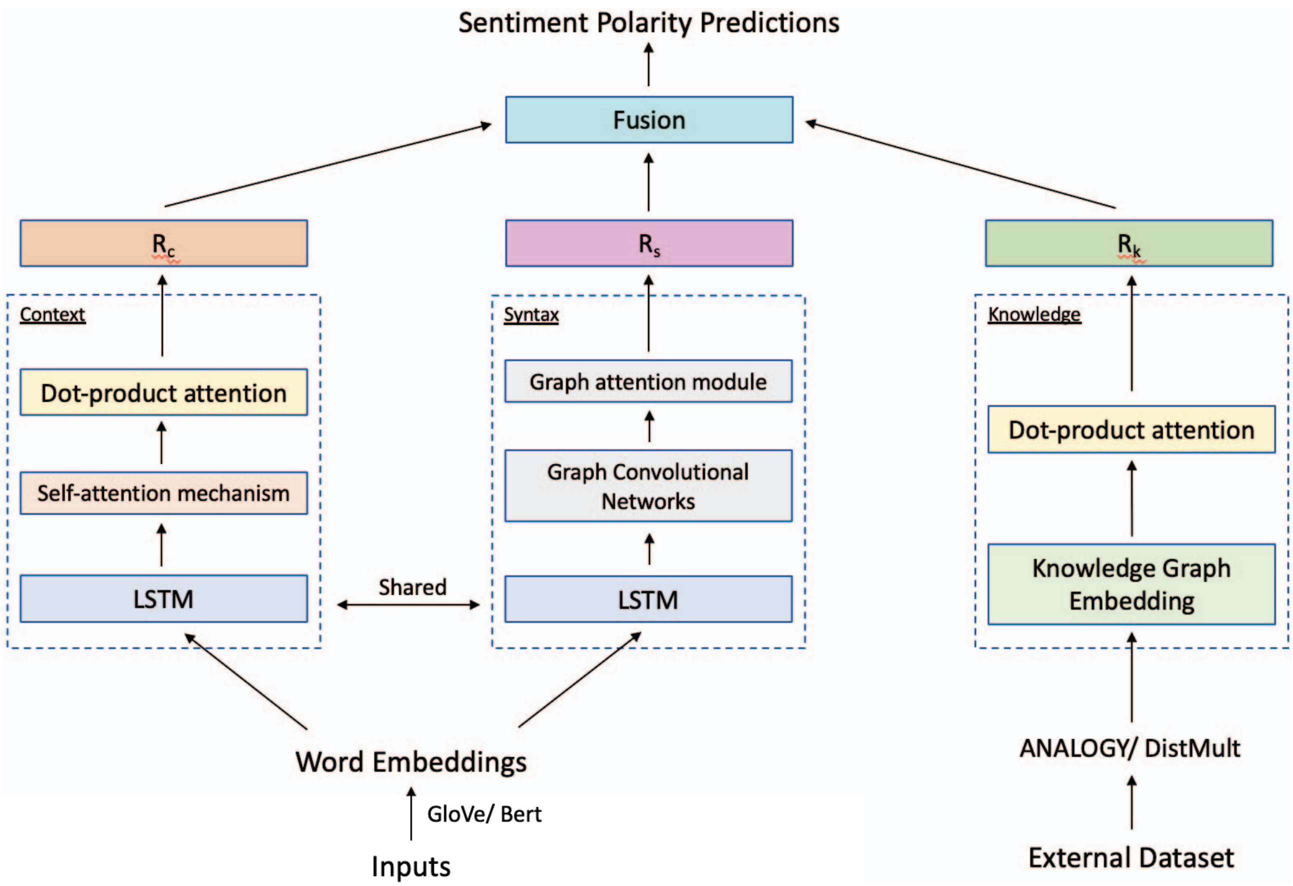

Fig. 1. Architecture overview of proposed method

We used Spacy[1] to construct the syntactic dependency tree and obtain the adjacency matrix of each word. Similar to the results produced by Kipf and Welling [75] and Yao et al. [76], we found that a 2-layer GCN was optimal as it outperformed a single layer GCN but the additional increase of layers did not increase the performance. Furthermore, we employed a graph-based attention mechanism to learn syntax-aware representations of each aspect by masking the non-aspect words with zeroes. We then used a dot-product attention mechanism to syntactically capture the relationships of the aspects and respective opinions.

C. Knowledge Section

For our external knowledge base, we utilized WordNet [77]. WordNet is a lexical database containing over 166,000 pairs of word forms and sense pairs as well as the semantic relations between word senses such as synonymy, antonymy, hyponymy, meronymy, troponymy and entailment [77]. This enables the model to understand the meaning of uncommon words by using more common and understandable words. For example, "parathas" is a flatbread which can be classified under the "food" or "bread" aspects, reducing the difficulty of comprehending the sentence. Rather than directly using the graph structure of the knowledge base, which was done by Zhou et al. [7], we instead used semantic matching approaches, such as ANALOGY [9], [60] and DistMult [9], [62], to model the semantic relations into knowledge graph embeddings. To reduce noise, we concatenated the embeddings with hidden state vectors. An attention mechanism is also leveraged to capture relevant semantic features as knowledge representations for the various aspects.

[1] https://spacy.io

D. Fusion

From the Context, Knowledge, and Syntax sections, we obtain the respective representations R_c, R_k, R_s. To fully utilize the representations, we first concatenate the representations in a pairwise manner, i.e., $R_c + R_s; R_c + R_k; R_k + R_s$ to obtain sentiment features R_{cs}, R_{ck}, R_{ks} respectively. We then feed these features into a three-by-three convolution layer to obtain the predicted sentiment which is ultimately used in the loss function to train and optimize the model. This 2-step fusion process improves the integration of external knowledge with contextual and syntactic knowledge, thus achieving greater performance.

IV. Experiment, Results and Discussions

A. Dataset and Experiment Setting

Our experiments were conducted using Restaurant14 and Laptop14 datasets obtained from the 2014 SemEval2014 ABSA Task 4 Subtask 2 challenge [78]. These two public datasets are widely employed by many researchers to evaluate sentiment analysis performance. Similar to Tang et al. [79], we removed instances of data where conflict occurred. A conflict occurs when an aspect term has both positive and negative sentiments. For example, the sentence "I wouldn't mind going back some time, but not before trying others nearby" contains both positive and negative polarities for the same aspect. Unlike Zhong et al. [9], we modified the processing of the dataset in different ways. Specifically, we changed how the distance is calculated between each word token and the sentiment by taking the shortest distance of each term to the left-most or right-most polarity term. This leverages on the idea that tokens closer to the aspect are of greater importance and usefulness, utilizing the positional information of the tokens in the sentence in the training of the model.

B. Results

In order to demonstrate the effectiveness of the proposed algorithm in this paper, the accuracy and F1 have been chosen as performance evaluation indicators [80], [81], our results in table I demonstrated the efficacy of our modifications in the dataset processing and model architecture as these improvements resulted in an improvement of 1.43%. As observed in Table I, the modification in dataset processing led to an increase in accuracy and we were able to further improve on it through various improvements in the model's architecture such as using a different optimizer for gradient descent or activation function.

TABLE I
PERFORMANCE OF THE PROPOSED MODEL VERSUS EXISTING METHODS

Model	Datasets			
	Restaurant14		Laptop14	
	Accuracy	*F1*	*Accuracy*	*F1*
Sentic-LSTM [31]	79.43	70.32	70.88	67.19
MTKEN [82]	79.47	68.08	73.43	69.12
SK-GCN [7]	81.53	72.90	77.62	73.84
KGAN [9]	84.46	**77.47**	78.91	75.21
Our Model	**85.09**	76.03	**80.34**	**75.36**

C. Discussion

Incorporating external knowledge sources, such as KGs, has been demonstrated as a feasible approach, as indicated by the results in Table I. In our research, we explored two distinct approaches to enhance the performance of the proposed method: model architecture and dataset processing improvements.

1) Model Architecture Improvement:

About model architecture improvement, the proposed model incorporates critical components, including an Optimizer, Learning Rate Scheduler, Weight Initializer, and Non-Linear Activation Function, to augment the performance of the ABSA task.

i) Optimizer

To improve the model architecture, we evaluated three different gradient descent optimization algorithms: Adam [83], AdamW [84] and Stochastic Gradient Descent (SGD) [85]. Similar to Loshchilov and Hutter [84], AdamW facilitated an improvement in the model's performance across both datasets but SGD did not. AdamW corrects the incorrect weight decay implemented in Adam, and this minor change in algorithm could explain the improved performance observed. In our experiments, we found that SGD performed the worst and we believe that a possible reason for this could be the use of attention mechanisms in our model as Zhang et al. [86] demonstrated that SGD does not perform as well as Adam in attention models.

ii) Learning Rate scheduler

Learning rate warmup is a common technique used to improve the performance of DNN in various domains such as computer vision [87] and natural language processing [88]. A learning rate warmup employs a smaller learning rate in the initial stages of the model's training. The intuition behind using a smaller learning rate at the start of the training is to overcome training instability as the model may observe and learn strongly-featured observations during the initial training which may cause it to skew towards these features. However, by using a smaller learn rate, the learning effect is reduced, enabling the model to learn the important features, improving the model's performance [87].

While prior research [89] suggests that a learning rate warmup can enhance a model's performance, our experiments, including schedulers like Linear, Exponential, One Cycle [90], and Stochastic Gradient Descent with Warm Restarts [91], did not yield positive results. Despite various attempts, none of these schedulers improved the model's performance. Interestingly, our findings align with Gotmare et al. [92], indicating that the linear scheduler tends to produce the best results.

We hypothesize a possible reason why the introduction of a learning rate warmup reduced the model's performance could be due to the fact the dataset is not highly differentiated, and that features observed in the initial training were salient to the model's learning process, thus, the reduced initial learning rate actually harmed rather than helped the model's learning process.

iii) Weight Initializer

Glorot and Bengio [93] have shown that random initialization of weights of the hidden layers of DNN often result in poor performance and thus they proposed a new initialization method, known as Glorot or Xavier initialization, which initializes the weight based on the number of inputs (fan-in) and outputs (fan-out) to a layer to improve model performance. However, Kaiming et al. [94] proposed a new initialization technique that is better suited for DNN containing asymmetric, non-linear activation functions, such as Rectified Linear Unit (ReLU).

Our experiments showed that Xavier initialization slightly outperformed Kaiming initialization in terms of accuracy, however, Kaiming initialization was able to achieve a higher F1 score. Interestingly, using Leaky ReLU as the non-linearity function, we were able to achieve a slightly better performance as compared to ReLU. We believe that this could possibly abe ttributed to the ReLU units dying during training. The dying ReLU problem [95] can occur during the initialization of the neurons whereby poor initialization leads to some neurons arriving at large negative values from which they are unable to recover. This causes the ReLU units to only output 0 for any input, affecting backpropagation and ultimately, the model's performance. Leaky ReLU attempts to address this problem by implementing a non-zero gradient to enable the neuron to recover, thus avoiding the dying ReLU problem.

iv) Non-Linear Activation Function

As Leaky ReLU seemed to perform better than ReLU, we hypothesized that the model suffers from a dying ReLU problem, thus, we implemented Parametric Rectified Linear Unit (PReLU) instead of ReLU as the activation function. Similar to leaky ReLU, PReLU does not assign a zero value for negative inputs but unlike leaky ReLU, which uses a predefined value, α, to multiple the negative values, PReLU treats α as a learnable parameter which is learnt during the model's training. Our results showed that by using PReLU instead of ReLU, we were able to further improve on the model's performance, strengthening the hypothesis that the model suffers from a dying ReLU problem.

2) Dataset Processing Improvement:

Regarding the processing of the dataset, we improved on the distance calculation between the various word tokens and the sentiment word as well as the weight applied to each word token. This change in distance calculation affects the weight applied to each token which ultimately influences the model's performance and our results showed that this improvement in data processing did improve the performance. The weight assigned to each token should be thought of as a type of hyperparameter as our experiments showed that changing the weight only may not necessarily improve the performance. On its own, the data processing step may not necessarily improve the model's performance but when this change is implemented alongside the change in model architecture, specifically the use of PReLU as the non-linear activation function, we achieved an improvement in accuracy and/or F1 over Zhong et al. [9]. By using AdamW, we can further improve the accuracy but suffer a slight decrease in F1 score for the Laptop14 dataset. Our experiments have demonstrated this change in dataset processing yielded an increase in performance across both datasets. As dataset processing is model agnostic, any current or future work that implements our modified dataset processing step can potentially gain an increase in performance. As the dataset processing used by Zhong et al. [9] was itself borrowed from Tang et al. [79], we believe that any future works extending from either author may achieve even better performance if the aforementioned change in dataset processing is implemented.

V. Conclusion

In this paper, we extended upon the work of existing ABSA to produce a new, less computationally complex model that has the capability to achieve a higher accuracy and F1 score. By leveraging knowledge graphs to enrich ABSA, the proposed method achieved a greater performance on both Laptop14 and Restaurant14 datasets.

This research demonstrates that changes to the dataset processing and model architecture can potentially improve the performance of models for ABSA tasks. Looking ahead, this work opens new avenues for research and development in sentiment analysis. As sentiment analysis continues to evolve, the insights gained here lay the groundwork for further investigations, pushing the boundaries of our ability to extract valuable sentiment insights from textual data in various domains. Future work can explore the integration of more extensive external knowledge sources and innovative techniques to further refine ABSA models, enhancing their effectiveness and applicability in real-world scenarios.

Acknowledgment

This work received support from A*STAR under the Advanced Manufacturing and Engineering (AME) Programmatic Grant (Award No.: A19E2b0098) and the Natural Science Foundation of Tianjin (19JCQNJC06000). The authors appreciate the contributions of SMU student assistants Glen Low, Samuel, Ong Zhi Yi, and Bui Phuong Thao. Special thanks to Geng Minghong, an SMU PhD student, for assistance in configuring and running codes on the GPU/server infrastructure.

References

[1] J. Cui, Z. Wang, S.-B. Ho, and E. Cambria, "Survey on sentiment analysis: evolution of research methods and topics," *Artificial Intelligence Review*, vol. 56, pp. 8469–8510, 2023.

[2] Z. Wang, Z. Hu, F. Li, S.-B. Ho, and E. Cambria, "Learning-based stock trending prediction by incorporating technical indicators and social media sentiment," *Cognitive Computation*, vol. 15, pp. 1092–1102, 2023.

[3] Z. Hu, Z. Wang, Y. Wang, and A.-H. Tan, "Aspect sentiment triplet extraction incorporating syntactic constituency parsing tree and commonsense knowledge graph," *Cognitive Computation*, vol. 15, pp. 337–347, 2023.

[4] K. He, R. Mao, T. Gong, C. Li, and E. Cambria, "Meta-based self-training and re-weighting for aspect-based sentiment analysis," *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 1731–1742, 2022.

[5] M. Hoang, O. A. Bihorac, and J. Rouces, "Aspect-based sentiment analysis using bert," in *Proceedings of the 22nd nordic conference on computational linguistics*, 2019, pp. 187–196.

[6] M. Yang, W. Tu, J. Wang, F. Xu, and X. Chen, "Attention based lstm for target dependent sentiment classification," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2017.

[7] J. Zhou, J. X. Huang, Q. V. Hu, and L. He, "Sk-gcn: Modeling syntax and knowledge via graph convolutional network for aspect-level sentiment classification," *Knowledge-Based Systems*, vol. 205, p. 106292, 2020.

[8] A. Zhao and Y. Yu, "Knowledge-enabled bert for aspect-based sentiment analysis," *Knowledge-Based Systems*, vol. 227, p. 107220, 2021.

[9] Q. Zhong, L. Ding, J. Liu, B. Du, H. Jin, and D. Tao, "Knowledge graph augmented network towards multiview representation learning for aspect-based sentiment analysis," *IEEE Transactions on Knowledge and Data Engineering*, 2023.

[10] Z. Wang, S.-B. Ho, and E. Cambria, "A review of emotion sensing: categorization models and algorithms," *Multimedia Tools and Applications*, vol. 79, pp. 35 553–35 582, 2020.

[11] Z. Wang, V. Joo, C. Tong, X. Xin, and H. C. Chin, "Anomaly detection through enhanced sentiment analysis on social media data," in *2014 IEEE 6th international conference on cloud computing technology and science*. IEEE, 2014, pp. 917–922.

[12] Z. Wang, V. Joo, C. Tong, and D. Chan, "Issues of social data analytics with a new method for sentiment analysis of social media data," in *2014 IEEE 6th International Conference on Cloud Computing Technology and Science*. IEEE, 2014, pp. 899–904.

[13] E. Cambria, Q. Liu, S. Decherchi, F. Xing, and K. Kwok, "SenticNet 7: A commonsense-based neurosymbolic ai framework for explainable sentiment analysis," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 2022, pp. 3829–3839.

[14] Z. Wang, V. J. C. Tong, P. Ruan, and F. Li, "Lexicon knowledge extraction with sentiment polarity computation," in *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2016, pp. 978–983.

[15] L. Oneto, F. Bisio, E. Cambria, and D. Anguita, "Statistical learning theory and ELM for big social data analysis," *IEEE Computational Intelligence Magazine*, vol. 11, no. 3, pp. 45–55, 2016.

[16] E. Cambria, A. Hussain, C. Havasi, and C. Eckl, "Common sense computing: From the society of mind to digital intuition and beyond," in *Biometric ID Management and Multimodal Communication*, ser. Lecture Notes in Computer Science. Berlin Heidelberg: Springer, 2009, vol. 5707, pp. 252–259.

[17] F. Xing, F. Pallucchini, and E. Cambria, "Cognitive-inspired domain adaptation of sentiment lexicons," *Information Processing and Management*, vol. 56, no. 3, pp. 554–564, 2019.

[18] E. Cambria, T. Mazzocco, A. Hussain, and C. Eckl, "Sentic medoids: Organizing affective common sense knowledge in a multi-dimensional vector space," ser. Lecture Notes in Computer Science. Berlin Heidelberg: Springer-Verlag, 2011, vol. 6677, pp. 601–610.

[19] E. Cambria, B. Schuller, B. Liu, H. Wang, and C. Havasi, "Statistical approaches to concept-level sentiment analysis," *IEEE Intelligent Systems*, vol. 28, no. 3, pp. 6–9, 2013.

[20] A. Valdivia, V. Luzón, E. Cambria, and F. Herrera, "Consensus vote models for detecting and filtering neutrality in sentiment analysis," *Information Fusion*, vol. 44, pp. 126–135, 2018.

[21] E. Cambria, R. Mao, M. Chen, Z. Wang, and S.-B. Ho, "Seven pillars for the future of AI," *IEEE Intelligent Systems*, vol. 38, no. 6, 2023.

[22] H. T. Nguyen, P. H. Duong, and E. Cambria, "Learning short-text semantic similarity with word embeddings and external knowledge sources," *Knowledge-Based Systems*, vol. 182, no. 104842, 2019.

[23] D. Vilares, H. Peng, R. Satapathy, and E. Cambria, "BabelSenticNet: A commonsense reasoning framework for multilingual sentiment analysis," in *IEEE SSCI*, 2018, pp. 1292–1298.

[24] E. Cambria, N. Howard, J. Hsu, and A. Hussain, "Sentic blending: Scalable multimodal fusion for continuous interpretation of semantics and sentics," in *IEEE SSCI*, Singapore, 2013, pp. 108–117.

[25] S. Cavallari, E. Cambria, H. Cai, K. Chang, and V. Zheng, "Embedding both finite and infinite communities on graph," *IEEE Computational Intelligence Magazine*, vol. 14, no. 3, pp. 39–50, 2019.

[26] E. Cambria, P. Chandra, A. Sharma, and A. Hussain, "Do not feel the trolls," in *ISWC*, Shanghai, 2010.

[27] F. Xing, E. Cambria, and R. Welsch, "Intelligent asset allocation via market sentiment views," *IEEE Computational Intelligence Magazine*, vol. 13, no. 4, pp. 25–34, 2018.

[28] E. Cambria, T. Benson, C. Eckl, and A. Hussain, "Sentic PROMs: Application of sentic computing to the development of a novel unified framework for measuring health-care quality," *Expert Systems with Applications*, vol. 39, no. 12, pp. 10 533–10 543, 2012.

[29] Z. Wang, V. J. C. Tong, and H. C. Chin, "Enhancing machine-learning methods for sentiment classification of web data," in *Asia Information Retrieval Societies Conference*. Springer, 2014, pp. 394–405.

[30] D. Rajagopal, E. Cambria, D. Olsher, and K. Kwok, "A graph-based approach to commonsense concept extraction and semantic similarity detection," in *WWW*, 2013, pp. 565–570.

[31] Y. Ma, H. Peng, T. Khan, E. Cambria, and A. Hussain, "Sentic LSTM: a hybrid network for targeted aspect-based sentiment analysis," *Cognitive Computation*, vol. 10, pp. 639–650, 2018.

[32] Z. Hu, Z. Wang, Y. Wang, and A.-H. Tan, "Msrl-net: A multi-level semantic relation-enhanced learning network for aspect-based sentiment analysis," *Expert Systems with Applications*, vol. 217, p. 119492, 2023.

[33] Z. Wang, Z. Hu, S.-B. Ho, E. Cambria, and A.-H. Tan, "MiMuSA—mimicking human language understanding for fine-grained multi-class sentiment analysis," *Neural Computing and Applications*, vol. 35, pp. 15 907–15 921, 2023.

[34] S. Poria, E. Cambria, and A. Gelbukh, "Aspect extraction for opinion mining with a deep convolutional neural network," *Knowledge-Based Systems*, vol. 108, pp. 42–49, 2016.

[35] Z. Toh and J. Su, "Nlangp at semeval-2016 task 5: Improving aspect based sentiment analysis using neural network features," in *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, 2016, pp. 282–288.

[36] L. Xu, J. Lin, L. Wang, C. Yin, and J. Wang, "Deep convolutional neural network based approach for aspect-based sentiment analysis," *Advanced Science and Technology Letters*, vol. 143, pp. 199–204, 2017.

[37] K. Sun, R. Zhang, S. Mensah, Y. Mao, and X. Liu, "Aspect-level sentiment analysis via convolution over dependency tree," 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)*, 2019, pp. 5679–5688.

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

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

[40] X. Li and W. Lam, "Deep multi-task learning for aspect term extraction with memory interaction," in *Proceedings of the 2017 conference on empirical methods in natural language processing*, 2017, pp. 2886–2892.

[41] X. Li, L. Bing, P. Li, W. Lam, and Z. Yang, "Aspect term extraction with history attention and selective transformation," *arXiv preprint arXiv:1805.00760*, 2018.

[42] P. Chen, B. Xu, M. Yang, and S. Li, "Clause sentiment identification based on convolutional neural network with context embedding," in *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*. IEEE, 2016, pp. 1532–1538.

[43] W. Xue, W. Zhou, T. Li, and Q. Wang, "Mtna: a neural multi-task model for aspect category classification and aspect term extraction on restaurant reviews," in *IJCNLP*, 2017, pp. 151–156.

[44] S. Liao, J. Wang, R. Yu, K. Sato, and Z. Cheng, "Cnn for situations understanding based on sentiment analysis of twitter data," *Procedia computer science*, vol. 111, pp. 376–381, 2017.

[45] W. Li, B. Jin, and Y. Quan, "Review of research on text sentiment analysis based on deep learning," *Open Access Library Journal*, vol. 7, no. 3, pp. 1–8, 2020.

[46] A. Severyn and A. Moschitti, "Twitter sentiment analysis with deep convolutional neural networks," in *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*, 2015, pp. 959–962.

[47] H. S. Sharaf Al-deen, Z. Zeng, R. Al-sabri, and A. Hekmat, "An improved model for analyzing textual sentiment based on a deep neural network using multi-head attention mechanism," *Applied System Innovation*, vol. 4, no. 4, p. 85, 2021.

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

[49] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based lstm for aspect-level sentiment classification," in *Proceedings of the 2016 conference on empirical methods in natural language processing*, 2016, pp. 606–615.

[50] S. Ji, S. Pan, E. Cambria, P. Marttinen, and S. Y. Philip, "A survey on knowledge graphs: Representation, acquisition, and applications," *IEEE*

transactions on neural networks and learning systems, vol. 33, no. 2, pp. 494–514, 2022.
[51] W. Li, L. Zhu, R. Mao, and E. Cambria, "SKIER: A symbolic knowledge integrated model for conversational emotion recognition," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023, pp. 13 121–13 129.
[52] Q. Wang, Z. Mao, B. Wang, and L. Guo, "Knowledge graph embedding: A survey of approaches and applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 12, pp. 2724–2743, 2017.
[53] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. Van Kleef, S. Auer *et al.*, "Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia," *Semantic web*, vol. 6, no. 2, pp. 167–195, 2015.
[54] M. Fabian, K. Gjergji, W. Gerhard *et al.*, "Yago: A core of semantic knowledge unifying wordnet and wikipedia," in *16th International world wide web conference, WWW*, 2007, pp. 697–706.
[55] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase: a collaboratively created graph database for structuring human knowledge," in *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, 2008, pp. 1247–1250.
[56] R. Hoffmann, C. Zhang, X. Ling, L. Zettlemoyer, and D. S. Weld, "Knowledge-based weak supervision for information extraction of overlapping relations," in *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*, 2011, pp. 541–550.
[57] A. Bordes, S. Chopra, and J. Weston, "Question answering with subgraph embeddings," *arXiv preprint arXiv:1406.3676*, 2014.
[58] J. Berant, A. Chou, R. Frostig, and P. Liang, "Semantic parsing on freebase from question-answer pairs," in *Proceedings of the 2013 conference on empirical methods in natural language processing*, 2013, pp. 1533–1544.
[59] M. Nickel, V. Tresp, H.-P. Kriegel *et al.*, "A three-way model for collective learning on multi-relational data." in *Icml*, vol. 11, 2011, pp. 3 104 482–3 104 584.
[60] H. Liu, Y. Wu, and Y. Yang, "Analogical inference for multi-relational embeddings," in *International conference on machine learning*. PMLR, 2017, pp. 2168–2178.
[61] B. Yang, W.-t. Yih, X. He, J. Gao, and L. Deng, "Embedding entities and relations for learning and inference in knowledge bases," *arXiv preprint arXiv:1412.6575*, 2014.
[62] M. Nickel, L. Rosasco, and T. Poggio, "Holographic embeddings of knowledge graphs," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 30, 2016.
[63] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, and G. Bouchard, "Complex embeddings for simple link prediction," in *International conference on machine learning*. PMLR, 2016, pp. 2071–2080.
[64] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," *Advances in neural information processing systems*, vol. 26, 2013.
[65] Z. Wang, J. Zhang, J. Feng, and Z. Chen, "Knowledge graph embedding by translating on hyperplanes," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 28, 2014.
[66] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, "Learning entity and relation embeddings for knowledge graph completion," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 29, 2015.
[67] G. Ji, S. He, L. Xu, K. Liu, and J. Zhao, "Knowledge graph embedding via dynamic mapping matrix," in *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: Long papers)*, 2015, pp. 687–696.
[68] M. Fan, Q. Zhou, E. Chang, and F. Zheng, "Transition-based knowledge graph embedding with relational mapping properties," in *Proceedings of the 28th Pacific Asia conference on language, information and computing*, 2014, pp. 328–337.
[69] J. Feng, M. Huang, M. Wang, M. Zhou, Y. Hao, and X. Zhu, "Knowledge graph embedding by flexible translation," in *Fifteenth International Conference on the Principles of Knowledge Representation and Reasoning*, 2016.
[70] H. Xiao, M. Huang, Y. Hao, and X. Zhu, "Transa: An adaptive approach for knowledge graph embedding," *arXiv preprint arXiv:1509.05490*, 2015.
[71] L. Ding, L. Wang, D. Wu, D. Tao, and Z. Tu, "Context-aware cross-attention for non-autoregressive translation," *arXiv preprint arXiv:2011.00770*, 2020.
[72] B. Yang, L. Wang, D. F. Wong, S. Shi, and Z. Tu, "Context-aware self-attention networks for natural language processing," *Neurocomputing*, vol. 458, pp. 157–169, 2021.
[73] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.
[74] D. Tang, B. Qin, and T. Liu, "Aspect level sentiment classification with deep memory network," *arXiv preprint arXiv:1605.08900*, 2016.
[75] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.
[76] L. Yao, C. Mao, and Y. Luo, "Graph convolutional networks for text classification," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, 2019, pp. 7370–7377.
[77] G. A. Miller, "Wordnet: a lexical database for english," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.
[78] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. AL-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq *et al.*, "Semeval-2016 task 5: Aspect based sentiment analysis," in *ProWorkshop on Semantic Evaluation (SemEval-2016)*. Association for Computational Linguistics, 2016, pp. 19–30.
[79] J. Tang, Z. Lu, J. Su, Y. Ge, L. Song, L. Sun, and J. Luo, "Progressive self-supervised attention learning for aspect-level sentiment analysis," *arXiv preprint arXiv:1906.01213*, 2019.
[80] Z. Wang and Z. Lin, "Optimal feature selection for learning-based algorithms for sentiment classification," *Cognitive Computation*, vol. 12, pp. 238–248, 2020.
[81] Z. Wang, S.-B. Ho, and E. Cambria, "Multi-level fine-scaled sentiment sensing with ambivalence handling," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 28, no. 4, pp. 683–697, 2020.
[82] S. Wu, Y. Xu, F. Wu, Z. Yuan, Y. Huang, and X. Li, "Aspect-based sentiment analysis via fusing multiple sources of textual knowledge," *Knowledge-Based Systems*, vol. 183, p. 104868, 2019.
[83] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
[84] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.
[85] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," *The Annals of Mathematical Statistics*, pp. 462–466, 1952.
[86] J. Zhang, S. P. Karimireddy, A. Veit, S. Kim, S. Reddi, S. Kumar, and S. Sra, "Why are adaptive methods good for attention models?" in *Advances in Neural Information Processing Systems*, vol. 33, 2020.
[87] P. Goyal, P. Dollár, R. Girshick, P. Noordhuis, L. Wesolowski, A. Kyrola, A. Tulloch, Y. Jia, and K. He, "Accurate, large minibatch sgd: Training imagenet in 1 hour," *arXiv preprint arXiv:1706.02677*, 2018.
[88] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, vol. 30, 2017.
[89] E. Hoffer, I. Hubara, and D. Soudry, "Train longer, generalize better: closing the generalization gap in large batch training of neural networks," in *Advances in neural information processing systems*, vol. 30, 2017.
[90] L. N. Smith and N. Topin, "Super-convergence: Very fast training of neural networks using large learning rates," in *Artificial intelligence and machine learning for multi-domain operations applications*, vol. 11006. SPIE, 2019, pp. 369–386.
[91] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," *arXiv preprint arXiv:1608.03983*, 2017.
[92] A. Gotmare, N. S. Keskar, C. Xiong, and R. Socher, "A closer look at deep learning heuristics: Learning rate restarts, warmup and distillation," *arXiv preprint arXiv:1810.13243*, 2018.
[93] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 2010, pp. 249–256.
[94] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.
[95] L. Lu, Y. Shin, Y. Su, and G. E. Karniadakis, "Dying relu and initialization: Theory and numerical examples," *arXiv preprint arXiv:1903.06733*, 2019.