Aspect-based Sentiment Analysis using BERT with Disentangled Attention

Emanuel H. Silva [1] **Ricardo M. Marcacini** [1]

Abstract

Aspect-Based Sentiment Analysis (ABSA) tasks aim to identify consumers' opinions about different aspects of products or services. BERT-based language models have been used successfully in applications that require a deep understanding of the language, such as sentiment analysis. This paper investigates the use of disentangled learning to improve BERT-based textual representations in ABSA tasks. Motivated by the success of disentangled representation learning in the field of computer vision, which aims to obtain explanatory factors of the data representations, we explored the recent DeBERTa model (Decoding-enhanced BERT with Disentangled Attention) to disentangle the syntactic and semantics features from a BERT architecture. Experimental results show that incorporating disentangled attention and a simple fine-tuning strategy for downstream tasks outperforms state-of-the-art models in ABSA's benchmark datasets.

1. Introduction

Neural language models have enabled significant performance improvements in various natural language processing tasks. In particular, deep neural language models are currently referred to as state-of-the-art for aspect-based sentiment analysis (ABSA), where we aim to compute the sentiment of the consumer's textual opinion considering the aspects of a product or service (Dang et al., 2020). While traditional methods explore the general sentiment of an opinion, ABSA is a more challenging task as an opinion can contain several aspects (Santos et al., 2021; Marcacini et al., 2018). For example, in the text "*the smartphone's camera is great, but the screen is very small*", we have a positive sentiment for the "camera" aspect and negative sentiment for the "screen" aspect.

[1]Institute of Mathematics and Computer Sciences, University of São Paulo, Brazil. Correspondence to: Emanuel H. Silva <emanuel.huber@usp.br>.

Proceedings of the LatinX in AI (LXAI) Research workshop at ICML 2021.

BERT-based deep neural language models are widely used for ABSA (Song et al., 2019; Zeng et al., 2019; Rietzler et al., 2020; Karimi et al., 2020b;a). BERT explores the recent bidirectional encoder representations from Transformers, which processes texts considering the context of a word from both the left and right sides (Devlin et al., 2019). Thus, BERT can generate more semantic textual representations, where each word is mapped to an embedding vector that depends on the context in which it occurs in a sentence. In addition, BERT is a pre-trained language model in a large textual corpus and allows fine-tuning for a supervised downstream task such as ABSA. In fact, the vast majority of ABSA's recent work explores different BERT fine-tuning strategies.

In 2020, He et al. (2020) introduced a new architecture for BERT-based language model training called Decoding-enhanced BERT with Disentangled Attention (DeBERTa). Disentangled representation learning (DRL) is a popular concept in computer vision (Locatello et al., 2019). In disentangled representation, ideally, each neuron learns a complete feature independent of other neurons, and thus real-world objects can be represented by explanatory factors. If the representations are disentangled, then one could interpret the meaning of each element in the representation space (Bengio et al., 2013). In the textual data, the DRL has not yet been extensively explored as in computer vision, since the explanatory factors on image data are more intuitively defined as object rotations, translations, scale, etc. DeBERTa incorporates DRL concepts into language models through a disentangled attention mechanism He et al. (2020). While in the original BERT each word is represented by a vector that is the sum of its content embedding vector and its position embedding vector, DeBERTa explores a disentangled attention mechanism for computing four attention scores between words: content-to-content, content-to-position, position-to-content, and position-to-position — thereby disentangling semantic (content) and syntactic (position) representation of the textual data.

In this paper, we present a method for Aspect-Based Sentiment Analysis using BERT with Disentangled Attention (ABSA-DeBERTa). Our method is motivated by the observation that words representing aspects and sentiments have positional dependence in opinion texts, where they are usually close to each other. Moreover, the same aspect can be described in different ways by consumers (e.g., screen and display), i.e., there is a semantic dependency between aspects. While existing studies attempt to incorporate these characteristics through complex BERT fine-tuning strategies, we argue that ABSA-DeBERTa's disentangled attention mechanism naturally incorporates such dependencies between aspects and sentiments words. Thus, a simple model fine-tuning for the downstream task already obtains state-of-the-art results, as shown in the experimental results using benchmark datasets.

2. Disentangled Representation Learning

The disentangled representation learning aims to create data representations where each factor of representation means a unique explanatory factor of the data, turning the generated representations more human-interpretable (Bengio et al., 2013). Recent works on the natural language processing field brought this concept that is mainly studied in the computer vision field.

Explanatory factors for image data can be defined as visual concepts that change on the training data, features like object translation, rotation, color and shape change, and other aspects (Locatello et al., 2019). On text data, the task is more complex, since there is no consensus among researchers about what are the explanatory factors of a text that can lead to better predictions. Recent work defined the disentanglement for textual data in two major aspects: the style space and the content space (John et al., 2019). The style space represents writing styling features and the content space is the manifold where the semantic meaning of the data is explored. For example, the style embedding represents the text personality (or write styling of the author) and the content embedding refers to the semantic meaning of a sentence.

Some tasks may use both or only one of those spaces to represent the data. In emotion recognition tasks, the writing style can help the model to get better results (Wu & Jiang, 2019). Other studies (Bao et al., 2019; Chen et al., 2019) define the exploratory features on semantic and syntactic spaces and that disentangling syntactic features can help on the automatic text generation tasks.

The disentangled representation learning (DRL) on textual data has been explored on problems such as text style transfer (John et al., 2019), conditional text generation (Cheng et al., 2020), and sentiment analysis (Pergola et al., 2020). To the best of our knowledge, there is no known DRL method applied for Aspect-Based Sentiment Analysis (ABSA), where disentanglement can be used to improve the semantic and syntactic relationships between aspects and their respective sentiments.

Neural language models like Devlin et al. (2019) capture semantic and syntactic information, but they do not perform this representation explicitly, and therefore they are entangled. DeBERTa (He et al., 2020) proposes a Transformer-based model where position and content vectors are disentangled, different from the previously proposed approaches that added the position vector with the content vector. As discussed by Bao et al. (2019) and Chen et al. (2019), the separation into syntactic and semantic spaces is one of the forms of disentanglement in texts that can be beneficial for the final task.

3. ABSA-DeBERTa

An NLP task that can benefit from disentangled representations is the Aspect-Based Sentiment Analysis (ABSA), where the goal is to predict the sentiment of given aspects in a sentence. Since it is naturally necessary to focus on semantic and syntactic relationships between aspect and sentiment words, disentangled representations can be helpful.

The disentangled attention introduced by DeBERTa (He et al., 2020) proposes to separate the content and text position components. The main idea is to learn attention weights for each component, different from other proposals discussed above that sum the position vector to the content vector (Vaswani et al., 2017). This explicit separation allows the model to better split position and content components of the data, where position embedding is responsible for generating syntactic features and content embedding is responsible for semantic features.

$$\begin{aligned} A_{i,j} &= \{\boldsymbol{H_i}, \boldsymbol{P_{i|j}}\} \times \{\boldsymbol{H_j}, \boldsymbol{P_{j|i}}\}^{\mathrm{T}} \\ &= \boldsymbol{H_i}\boldsymbol{H_j}^{\top} + \boldsymbol{H_i}\boldsymbol{P_{j|i}}^{\top} + \boldsymbol{P}_{\boldsymbol{i}|j}\boldsymbol{H_j}^{\top} + \boldsymbol{P_{i|j}}\boldsymbol{P}_{j|i}^{\top} \end{aligned} \tag{1}$$

In BERT-based models, tokens are represented by content vectors and position vectors. In the case of DeBERTa, given two tokens i and j in a sentence, consider that $\boldsymbol{H_i}$ represents the content vector of token i and that $\boldsymbol{P_{i|j}}$ represents the relative position vector between token i and j. Equation 1 defines the cross-attention matrix used in DeBERTa. Note that the attention matrix $A_{i,j}$ stores an attention weight for each pair of words in a sentence. The final attention weight is calculated from four scores (disentangled matrices): content-to-content $\boldsymbol{H_i}\boldsymbol{H_j}^{\top}$, content-to-position $\boldsymbol{H_i}\boldsymbol{P_{j|i}}^{\top}$, position-to-content $\boldsymbol{P}_{\boldsymbol{i}|j}\boldsymbol{H_j}^{\top}$, and position-to-position $\boldsymbol{P_{i|j}}\boldsymbol{P}_{j|i}^{\top}$. He et al. (2020) show that the disentangled attention mechanism can be incorporated into BERT model without significant changes in the rest of the neural network architecture.

In ABSA-DeBERTa, the objective is to determine the sentiment of each aspect of a given entity. ABSA-DeBERTa takes advantage of DeBERTa's pre-trained model parameters to initialize the model for the downstream task, i.e., we apply fine-tuning to update model parameters according to ABSA labeled data. Unlike previous studies that propose complex fine-tuning strategies to incorporate syntactic and semantic characteristics of ABSA tasks, we follow BERT's original fine-tuning strategy since these characteristics are already naturally captured by the disentangled attention mechanism. During training, we consider the input (x, y), $x = (sentence[SEP]aspect[SEP])$, where $[SEP]$ is a special BERT token to separate two token sequences, and y is the sentiment label (e.g., positive, negative or neutral). We use the output of the language model to feed the downstream sentiment classification task (a standard softmax layer with categorical cross-entropy loss function). The classification error during training is used to adjust the initial parameters of DeBERTa model.

4. Experimental Evaluation

To evaluate the ABSA-DeBERTa, we used the Semeval benchmark datasets (Pontiki et al., 2014). There are two datasets for this task: the laptop and restaurants reviews. In these datasets, user evaluates their experience with a laptop or restaurant based on aspects and their respective qualities. The datasets contain 5.936 total samples and three sentiments (positive, negative, and neutral). Table 1 shows an overview of the datasets.

	Positive	Negative	Neutral
Laptops	994	870	464
Restaurants	2164	807	637

Table 1. Overview of the Semeval benchmark datasets.

We compared ABSA-DeBERTa with recent fine-tuning BERT strategies for ABSA tasks. BERT-SPC (Song et al., 2019) uses the original BERT model and classic fine-tuning strategy. LCF-BERT (Zeng et al., 2019) adds a local context during fine-tuning to emphasize the relative distance between tokens. BERT-ADA (Rietzler et al., 2020) explores different fine-tuning scenarios by incorporating texts from other sources and exploring cross-domain training. PH-SUM (Karimi et al., 2020b) uses different intermediate layers of the BERT model during fine-tuning. BAT (Karimi et al., 2020a) uses adversarial training to improve BERT's fine-tuning.

The ABSA-DeBERTa model architecture follows the same proposed by DeBERTa. In ABSA-DeBERTa fine-tuning, we use 32 epochs with a max sequence length of 256. The training data was randomly split with 80% for training and 20% for validation. The test set is provided by the Semeval competition and remained unchanged for final results calculations.

Table 2 shows an overview of the experimental results using ACC and F1 metrics for both Laptop and Restaurant datasets. Note that even using a simple fine-tuning strategy, our ABSA-DeBERTa outperforms recently proposed strategies for ABSA tasks, obtaining state-of-the-art results. We emphasize that the disentangled representation learning mechanism improves the performance of ABSA tasks, since the final textual representation contains the explanatory factors related to syntactic structures (position embedding) and semantics (content embedding).

5. Conclusion

This paper presents the use of disentangled attention mechanism for sentiment analysis, showing that the detachment of the position and content vectors can help to solve the ABSA tasks. The separation of position and content can be analogous to a certain extent to the separation of syntactic and semantic spaces, with the position only including the ordering feature of the text, but leaving other aspects of syntax aside. However, with the representation that uses only position and content spaces, it was possible to obtain initial experimental results that indicate promising results.

Directions for future work involve exploring the explanatory factors from disentangled attention mechanism for the final representation of the reviews. In other areas, such as computer vision, these factors are used to generate explainable models. The concepts of these areas can be used in a useful way to develop ABSA methods that present good classification performance (ACC and F1) together with model interpretation. ABSA-DeBERTa code is available at `https://github.com/huberemanuel/DeBERTa/`.

	Mean F1	F1 Laptop	F1 Restaurants	Mean Acc	Acc Laptop	Acc Restaurants
BERT-SPC (Song et al., 2019)	76.01	75.03	76.98	76.73	78.99	84.46
LCF-BERT (Zeng et al., 2019)	80.67	79.59	80.40	84.80	82.45	87.14
BERT-ADA (Rietzler et al., 2020)	77.78	75.77	79.79	83.81	79.94	87.69
PH-SUM (Karimi et al., 2020b)	78.10	76.52	79.67	82.89	79.40	86.37
BAT (Karimi et al., 2020a)	77.87	76.50	79.24	82.69	79.35	86.03
ABSA-DeBERTa (Ours)	81.39	79.36	83.42	86.11	82.76	89.46

Table 2. Experimental results comparison using Acc and Macro F1 scores

References

Bao, Y., Zhou, H., Huang, S., Li, L., Mou, L., Vechtomova, O., Dai, X., and Chen, J. Generating sentences from disentangled syntactic and semantic spaces. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 6008–6019, 2019.

Bengio, Y., Courville, A., and Vincent, P. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35 (8):1798–1828, jun 2013. ISSN 01628828.

Chen, M., Tang, Q., Wiseman, S., and Gimpel, K. A multi-task approach for disentangling syntax and semantics in sentence representations. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 2453–2464, 2019.

Cheng, P., Min, M. R., Shen, D., Malon, C., Zhang, Y., Li, Y., and Carin, L. Improving disentangled text representation learning with information-theoretic guidance. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 7530–7541, 2020.

Dang, N. C., Moreno-García, M. N., and De la Prieta, F. Sentiment analysis based on deep learning: A comparative study. *Electronics*, 9(3):483, 2020.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 4171–4186, 2019.

He, P., Liu, X., Gao, J., and Chen, W. DeBERTa: Decoding-enhanced BERT with Disentangled Attention. In: arXiv, 2020. URL https://arxiv.org/abs/2006.03654.

John, V., Mou, L., Bahuleyan, H., and Vechtomova, O. Disentangled representation learning for non-parallel text style transfer. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 424–434, 2019.

Karimi, A., Rossi, L., and Prati, A. Adversarial Training for Aspect-Based Sentiment Analysis with BERT. In: arXiv, 2020a. URL https://arxiv.org/abs/2001.11316.

Karimi, A., Rossi, L., and Prati, A. Improving BERT Performance for Aspect-Based Sentiment Analysis. In: arXiv., 2020b. URL https://arxiv.org/abs/2010.11731.

Locatello, F., Bauer, S., Lucic, M., Raetsch, G., Gelly, S., Schölkopf, B., and Bachem, O. Challenging common assumptions in the unsupervised learning of disentangled representations. In *International Conference on Machine Learning*, pp. 4114–4124, 2019.

Marcacini, R. M., Rossi, R. G., Matsuno, I. P., and Rezende, S. O. Cross-domain aspect extraction for sentiment analysis: A transductive learning approach. *Decision Support Systems*, 114:70–80, 2018.

Pergola, G., Gui, L., and He, Y. A Disentangled Adversarial Neural Topic Model for Separating Opinions from Plots in User Reviews. In: arXiv, 2020. URL https://arxiv.org/abs/2010.11384.

Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., and Manandhar, S. SemEval-2014 task 4: Aspect based sentiment analysis. In *International Workshop on Semantic Evaluation (SemEval 2014)*, pp. 27–35. ACL, August 2014.

Rietzler, A., Stabinger, S., Opitz, P., and Engl, S. Adapt or get left behind: Domain adaptation through bert language model finetuning for aspect-target sentiment classification. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pp. 4933–4941, 2020.

Santos, B. N., Marcacini, R. M., and Rezende, S. O. Multi-domain aspect extraction using bidirectional encoder representations from transformers. *IEEE Access*, 2021.

Song, Y., Wang, J., Jiang, T., Liu, Z., and Rao, Y. Targeted sentiment classification with attentional encoder network. In *International Conference on Artificial Neural Networks*, pp. 93–103. Springer, 2019.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. In *International Conference on Neural Information Processing Systems*, pp. 6000–6010, 2017.

Wu, Z. and Jiang, Y. Disentangling latent emotions of word embeddings on complex emotional narratives. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pp. 587–595. Springer, 2019.

Zeng, B., Yang, H., Xu, R., Zhou, W., and Han, X. Lcf: A local context focus mechanism for aspect-based sentiment classification. *Applied Sciences*, 9:3389, 08 2019.

Author contributions: E.H.R. and R.M.M. contributed to the design of the research. E.H.R. carried out the experiment. E.H.R. and R.M.M. wrote the manuscript. R.M.M. supervised the project.