Dimensional Aspect-Based Sentiment Analysis: DeBERTa-based Approach with Lexicon Integration

Anıl Dervişoğlu
AI & Data Engineering
Istanbul Techincal University
dervisoglua22@itu.edu.tr
150220344

Ömer Faruk Zeybek
AI & Data Engineering
Istanbul Technical University
zeybeko22@itu.edu.tr
150220743

I. Project Description

Our project focuses on "Dimensional Aspect-Based Sentiment Analysis (DimABSA)", task 3 in Semeval 2026. We decided to target subtask 1 (Dimensional Aspect Sentiment Rating - DimASR) to follow a focused and thorough approach.

The catch of this project is that, unlike traditional versions of sentiment analysis [3], we aren't solely classifying reviews with discrete labels of positive, negative or neutral. DimABSA focuses on predicting continuous emotional dimensions of valence and arousal between 1 and 9. This approach captures emotional level that categorical approaches can't represent, such as the difference between "good" (V=7.5, A=5.0) and "excellent" (V=8.5, A=7.0) which is directly taken from the competition page.

Our aim here is to develop a deep learning system based on "DeBERTa-V3" (which we found trough our initial methodology analysis) with specialized regression heads for dimensional prediction. We'll also integrate the NRC VAD Lexicon to improve predictions, especially for rare words. Per the guidelines, we've designed a labeling strategy to convert collected data into usable training samples. In the end we will move on in the order of: using the provided data, collecting and labeling additional data, implementing our main model along with multiple baselines to compare later on and finally conducting extensive ablation studies.

II. Problem Definition

For subtask 1, for each given aspect term a_i in sentence s, we predict continuous sentiment scores (v_i, ar_i) where:

- v_i represents valence, ranging between 1 (extremely negative) and 9 (extremely positive).
- ar_i represents arousal, ranging between 1 (very calm) and 9 (very excited).

Performance is measured using RMSE (used in evaluation):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[(v_i^{pred} - v_i^{gold})^2 + (ar_i^{pred} - ar_i^{gold})^2]} \tag{1}$$

We will calculate MAE and Pearson correlation as secondary metrics while using the following loss (to backprop):

$$L = MSE(v^{pred}, v^{gold}) + MSE(ar^{pred}, ar^{gold}) \tag{2}$$

III. Dataset

A. Initial Dataset

We will use the provided task dataset, which includes English customer reviews for restaurants and laptops with aspect terms and continuous valence and arousal scores annotated.

B. Data Collection and Labeling Strategy

In addition to the initial dataset we have to collect data as well per the guidelines. Our strategy is:

Web Scraping: We will extract restaurant and laptop reviews from Amazon and Yelp using Beautiful Soup and Selenium, collecting review text, star ratings (1-5), product or restaurant names and domain labels. We will then balance star rating distribution to avoid class imbalance.

Labeling: Scraped reviews don't come with aspect terms marked or continuous valence and arousal scores. To do so, we designed a 3 approached annotation pipeline:

Stage 1 - Finding aspects: We'll be using a hybrid extraction pipeline. First using spaCy library for Part-of-Speech (POS) tagging to find common nouns (like battery, service) and Named Entity Recognition (NER) to find branded products (MacBook, Whopper) with manuel control over samples.

Stage 2 - Converting Star Ratings to VA Scores (Approach A): For valence, we use star ratings as a guide: 5 star $\rightarrow$ $V = 8.0-9.0$, 4 star $\rightarrow V = 7.0-7.9$, 3 star $\rightarrow V = 5.0-6.9$, 2 star $\rightarrow V = 3.0 - 4.9$, 1 star $\rightarrow V = 1.0 - 2.9$

For arousal, star ratings don't help since calm and excited reviews can have the same rating. We will estimate arousal from exclamation marks, capitalization and strong adjectives from NRC VAD Lexicon values for words near the aspect.

Stage 2 Alternative - Aspect-Level Analysis (Approach B): Since star ratings apply to entire reviews while aspects can have varying sentiments, we will also:

- Use pre-trained ABSA models to get aspect scores, then convert categorical sentiment to valence
- Use NRC Affect Intensity lexicon alongside intensity words for arousal.

Stage 2 Alternative - LLM Annotation (Approach C): We will use GPT-4 (or similar cost effective model) for aspect VA annotation:

- First we will design few-shot prompts with 5-10 gold examples from Semeval data
- Request reasoning in prompt to improve consistency

All three approaches will be compared and potentially combined using weighted averaging based on confidence scores (calculation for alternatives: # aspect mentions, model confidence, llm confidence output).

Stage 3 - Model Assisted Annotation: We'll train an initial DeBERTa model on the official Semeval data, generate predictions on our collected data, and apply confidence filtering (uncertainty < 0.5 RMSE). We'll then manually review %10 of predictions, equally distrubed on domain and star rating.

Data Augmentation: We'll use back-translation (English to German to English), synonym replacement for non-aspects and aspect substitution with semantically similar aspects.

IV. Methodology

A. Our Main Model Architecture

We propose a DeBERTa-V3 [1] based framework with improved aspect context interaction:

Input Encoding: We will encode each training instance using the format [CLS] sentence [SEP] aspect [SEP], this allows model to focus on the relevant parts of the text.

Aspect Representation: For aspect representation we'll follow these steps:

- We'll compute attention weights over the aspect related tokens to find which parts have more importance with following formula $\alpha = \mathrm{softmax}(W_a \cdot [h_i; h_{aspect}])$
- Using these attention scores, we'll perform weighted pooling on aspect tokens and give higher weight to tokens that contribute more to the meaning of the aspect.
- For multi word aspects, we'll combine subword embeddings using learned weights and it'll enable model to capture nuances within multi token aspect expressions.

Aspect Context Interaction: We'll add a cross-attention layer so the aspect can focus to the context tokens most relevant for predicting its valence and arousal

Regression Heads: Two separate heads for V and A.

Output Scaling: Our primary approach uses $output = 1 + 8 \times \sigma(logit)$ to map logits between 1-9 range.

B. Lexicon Integration

We integrate the "NRC VAD Lexicon" [2] (55,000+ words with VA scores). Instead of a fixed ± 5 token window, we use dependency parsing to identify syntactically related words. We compute an 8-dimensional feature vector: (context_v, context_a, aspect_v, aspect_a, weighted_v, weighted_a, max_v, max_a), concatenate with DeBERTa representation ($h_{combined} = [h_{aspect}; f_{lexicon}]$) and feed to regression heads.

C. Training Details (will be experimented)

- Optimizer: Adam optimizer with different learning rates
- Loss: MSE for both valence and arousal, summed
- Training: Up to 20 epochs, early stopping, batch size 16 with gradient accumulation
- Regularization: Dropout 0.1, gradient clipping (norm 1)

V. Baseline Models

A. Baseline 1: Lexicon Only Prediction

Uses only the NRC VAD Lexicon without neural models, calculating the distance weighted average of VA scores for all words, blending context (0.3) and aspect scores (0.7). The purpose is to test if deep learning is necessary.

B. Baseline 2: Alternative Transformer Encoders

Compares DeBERTa against BERT-base [5] and RoBERTa-base [6] with identical regression heads and training. The purpose is to test if DeBERTa provides benefits.

C. Baseline 3: Few-Shot Large Language Models

Tests GPT-3.5 and GPT-4 in few-shot setting (as explained in data collection section) with prompt similar to: *"Task: Predict valence (1-9) and arousal (1-9) for the given aspect. [Examples] Sentence: {text} Aspect: {aspect} Output (V#A):"* The purpose is to test if LLM competes with our fine tuning.

D. Baseline 4: Simplified Fine-tuned Model

DeBERTa-V3 with single-layer heads and no lexicon integration. The purpose is to test gains from our complex architecture (two-layer heads + lexicon).

VI. Ablation Studies

Ablation 1- Training Details: Will be experimented with different lr values, optimizers such as SGD, Adabelief etc., epoches, batch sizes, regularization parameters.

Ablation 2- Lexicon Features: With vs. without NRC VAD integration. To test if lexicon helps or adds noise.

Ablation 3- Aspect Representation: Compare: (a) Mean pooling, (b) [CLS] token, (c) Attention-weighted pooling.

Ablation 4- Data Augmentation: Original vs back translation vs synonym replacement vs aspect substitution vs all.

Ablation 5- Output Scaling Functions: (a) Sigmoid: $1 + 8 \times \sigma(logit)$, our current approach (b) Tanh: $1 + 8 \times (\tanh(logit/2) + 1)/2$, to test if sigmoid makes it harder to score in boundaries (1 and 9) (c) Linear clipping: $\min(\max(logit, 1), 9)$, testing linear approach.

Ablation 6- Loss Functions: MSE vs. Huber loss vs. MAE vs. weighted combination. To test robustness to outliers.

VII. Conclusion

Our project focuses on an approach to the DimABSA subtask 1 challenge with DeBERTa-V3 and lexicon integration. We came up with four strong baselines (lexicon only, alternative encoders, few-shot LLMs, simplified fine tuning) and ablation studies covering architectural choices, training strategies and output formuls.

More importantly, we've designed a 3 way annotation pipeline that addresses the aspect level sentiment challenge, comparing star based, aspect focused and LLM based annotation strategies. We tried to utilize aspect representation with attention mechanisms and cross-attention for aspect context interaction. Depending on our results, we will apply more ablation tests as well, hoping to create a strong method.

References

[1] Y. He et al., "DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing," *arXiv preprint arXiv:2111.09543*, 2021.

[2] S. Mohammad, "Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words," in *Proc. 56th Annual Meeting of the Association for Computational Linguistics*, 2018, pp. 174–184.

[3] W. Xue and T. Li, "Aspect Based Sentiment Analysis with Gated Convolutional Networks," in *Proc. 56th Annual Meeting of the Association for Computational Linguistics*, 2018, pp. 2514–2523.

[4] Semeval-2026 Task 3, "Dimensional Aspect-Based Sentiment Analysis," *https://github.com/DimABSA/DimABSA2026*, 2025.

[5] J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.

[6] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," *arXiv preprint arXiv:1907.11692*, 2019.