**Methodology:**

The exploration into sentiment analysis of Amazon product reviews commenced with an investigation centered on text-based evaluations, employing the BERT (Bidirectional Encoder Representations from Transformers) model. This initial phase laid the groundwork for understanding the intrinsic value of textual content in expressing sentiment, prior to integrating more complex, multi-modal features.

**Experiment 1: Text-Based Sentiment Analysis Using BERT**

The foundation of this analysis was the application of a transformer-based neural network, specifically the BERT model, known for its proficiency in natural language understanding tasks. The architecture of the sentiment classifier was constructed upon a pre-trained BERT model, tailored to classify sentiments within Amazon product reviews into one of five predefined categories.

The dataset comprised a collection of product reviews, each annotated with a sentiment label. To prepare the data for the model, a custom dataset class, referred to as **AmazonDataset**, was developed. This class was responsible for processing the review texts, ensuring they conformed to the input requirements of the BERT model. Key preprocessing steps involved adjusting the length of each review to a fixed maximum, generating attention masks to signify the presence of actual content versus padding, and converting the text into a format understandable by the model, namely token ids.

The sentiment classifier extended the BERT architecture by incorporating a dropout layer, which served to mitigate overfitting, and a linear layer that mapped the high-dimensional output of BERT to the sentiment classes. The final output, termed **logits**, represented the probability distribution across the sentiment categories.

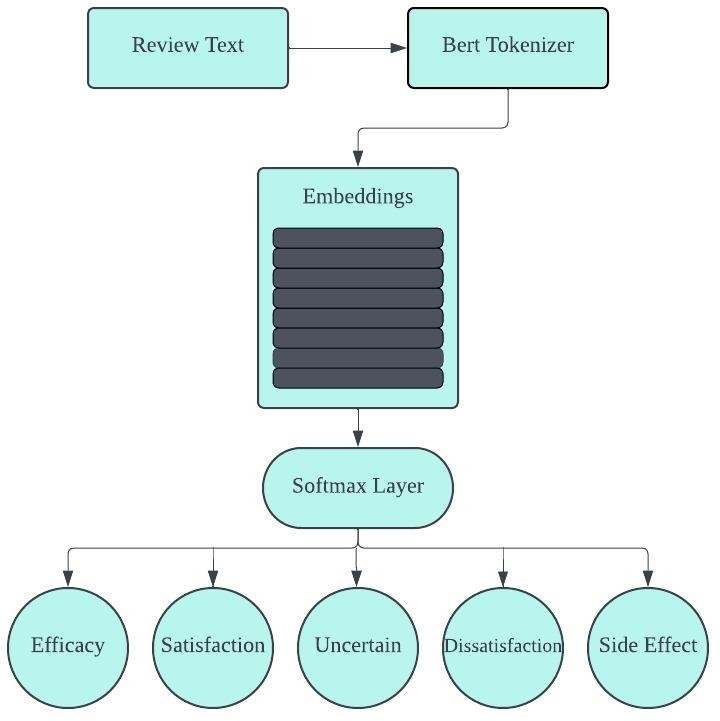
The training process entailed iterating over the dataset in batches, where each batch contained a set of tokenized review texts and their corresponding sentiment labels. The model's performance was gauged using cross-entropy loss, and optimization was conducted using the Adam algorithm, renowned for its efficiency in handling sparse gradients and adaptive learning rates.

This experiment was underpinned by the hypothesis that textual content alone can offer significant insights into the sentiment of product reviews. BERT's design, leveraging bidirectional context for token representations, provided a robust framework for capturing the nuanced sentiment expressed in the reviews. The choice of BERT was strategic, leveraging its pre-trained knowledge base to enhance the model's understanding of linguistic nuances.

The methodology adopted in this phase was instrumental in setting a baseline for sentiment analysis based solely on textual information. It illuminated the strengths of leveraging advanced NLP techniques for sentiment classification and identified potential areas where textual analysis might fall short, such as in capturing sentiments conveyed through non-verbal means like emojis.

Top of Form

[**Fig. 1**](#_bookmark2) Only Review Text BERT Model shows the architecture for this approach.



**Fig. 1.** Only fine-tuned BERT on the Review Text

The text-based sentiment analysis provided an essential baseline for assessing sentiment in e-commerce reviews, emphasizing the critical role of linguistic content. By employing a sophisticated model like BERT, this phase aimed to capture the depth of sentiment expressions in text form, setting a foundation for further explorations into multi-modal sentiment analysis. This methodological approach highlighted the potential and limitations of relying exclusively on text, paving the way for subsequent analyses that would incorporate additional features to enrich the sentiment analysis framework.

Building on the foundational insights from the initial text-based sentiment analysis, the second experiment expanded the scope to include emojis alongside review text. This approach aimed to explore how the addition of emojis, as a form of non-verbal communication, enhances the model's ability to interpret and classify sentiments more accurately.

**Experiment 2: Text with Emoji Model Using BERT**

Extended Methodological Framework

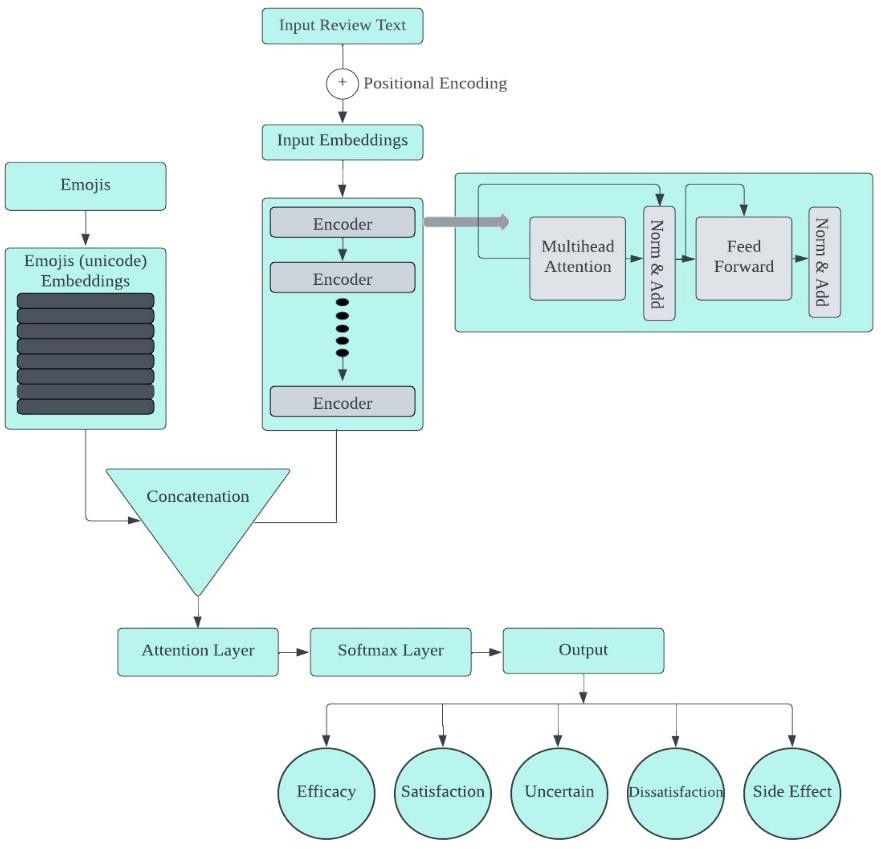
The integration of emojis into the sentiment analysis model introduced a multi-modal dimension to the data processing and analysis pipeline. The **AmazonDataset** class was adapted to accommodate not only the textual content of reviews but also the associated emojis. Each emoji within a review was converted to its textual representation using a process known as "demojization," which translates graphical emojis into their descriptive textual counterparts.

This process facilitated the inclusion of emojis as part of the input to the BERT model, allowing the model to process emojis in conjunction with textual data. The tokenizer was employed to encode both review text and emoji text, generating token ids, attention masks, and token type ids, ensuring that the model could distinguish between textual and emoji inputs.

The **SentimentClassifier** was enhanced to address the complexity of handling both text and emojis. The model incorporated separate attention mechanisms for text and emojis, enabling it to focus on salient features within both modalities. The text attention mechanism processed the output of the BERT model, generating attention scores for textual content. Simultaneously, an emoji attention mechanism was introduced, designed to map the emoji inputs to a higher-dimensional space aligned with the model's hidden size, facilitating the generation of attention scores for emoji content.

A critical component of this extended architecture was the integration of text and emoji representations. Attention scores were used to compute weighted sums of text and emoji features, which were then concatenated to form a combined representation. This combined representation was subsequently fed into a dense layer for classification, outputting logits that correspond to the probability distribution over sentiment classes.

Theoretical Considerations

The hypothesis driving this experiment was that the integration of emojis would provide additional context and emotional nuance, enhancing the model's ability to discern sentiment with greater accuracy. Emojis often convey subtle emotional undertones and sentiments that might not be explicitly stated in text, offering a complementary dimension to textual analysis.

**Fig. 2.** Model Architecture of Review Text with Custom Emoji Embeddings

This experiment leveraged the representational power of BERT for textual content while introducing an innovative approach to incorporating emoji information. The dual attention mechanism underscored the model's capacity to dynamically weigh textual and emoji inputs, reflecting the intuitive process humans employ when interpreting combined textual and visual cues.

**Conclusion**

The incorporation of emojis into the sentiment analysis model marked a significant methodological advancement, acknowledging the multifaceted nature of communication in e-commerce reviews. By extending the analysis to include emojis alongside text, the experiment aimed to capture a more holistic view of sentiment, reflecting the richness and complexity of consumer feedback. This approach not only built upon the foundational text-based model but also set the stage for further explorations into multi-modal sentiment analysis, promising to unveil deeper insights into consumer emotions and preferences.

Top of Form

The third experiment, forming the cornerstone of the sentiment analysis framework, introduced a comprehensive model that not only analyzed textual content but also integrated emojis, star ratings, and total votes as pivotal features. This multifaceted approach aimed to harness a broader spectrum of information contained within Amazon product reviews, promising a more nuanced understanding of consumer sentiments.

**Experiment 3: Multi-Feature BERT Model Incorporating Text, Emojis, Star Ratings, and Total Votes**

Advanced Methodological Framework

This experiment expanded upon the foundational models by embedding a richer set of features into the sentiment analysis process. The methodology was characterized by a series of sophisticated data processing and modeling steps tailored to accommodate the diverse nature of the input data.

The **AmazonDataset** class was enhanced to preprocess review texts, extract and convert emojis to embeddings, and include star ratings and total votes as part of the model input. Review texts underwent a thorough preprocessing routine to standardize and refine the textual content, including lowercasing, removing URLs, stripping punctuation, and applying tokenization and lemmatization to distill the text to its most informative components.

Emojis within reviews were treated with particular attention, being first demojized and then transformed into embeddings. A custom embedding layer was designed to represent emojis, with dimensions aligned with the BERT model's hidden size to ensure seamless integration into the model architecture.

The **AmazonBERTClassifier** model was a pivotal element of the methodology, embodying the multi-modal sentiment analysis approach. It featured a multihead self-attention mechanism applied to the output of the BERT model, enhancing the model's capacity to focus on relevant aspects of the textual content. An innovative emoji-aware attention mechanism was introduced, utilizing the emoji embeddings to inform the model of the emotional and contextual nuances conveyed by emojis within the text.

A fusion layer combined the outputs of the self-attention and emoji-aware attention mechanisms, integrating the insights drawn from both textual and emoji analyses. The fusion output then underwent average pooling to consolidate the information into a format suitable for sentiment classification.

Theoretical Considerations

This multi-feature model was grounded in the hypothesis that a comprehensive analysis of Amazon reviews, incorporating text, emojis, star ratings, and total votes, would offer a more complete picture of consumer sentiment. The integration of emojis, in particular, was anticipated to enrich the model's interpretative capabilities, given that emojis often encapsulate emotions and sentiments not explicitly expressed in text.

The model's architecture was carefully crafted to respect the distinct contributions of each feature type. The self-attention and emoji-aware attention mechanisms mirrored the cognitive process humans employ when interpreting complex communications, weighing textual content and emotional cues conveyed through emojis to form an understanding.

**Conclusion**

The multi-feature sentiment analysis model represented a significant leap forward in the field, embodying a holistic approach to understanding consumer sentiment on e-commerce platforms. By weaving together textual analysis with insights drawn from emojis, star ratings, and total votes, the model aimed to capture the multifaceted nature of consumer feedback, offering a richer, more nuanced view of sentiment than text-alone analyses could provide. This comprehensive methodology not only set a new standard for sentiment analysis in e-commerce but also highlighted the potential for multi-modal approaches in broader natural language processing applications.

Top of Form