Aggregated Token Attention Network (ATAN) for Food Review

Sentiment Analysis

Farjana Sultana Samia

IOWA STATE UNIVERSITY

Abstract

- Sentiment classification of Amazon Fine Food Reviews
- Pre-trained BERT models.
- Evaluates models accuracy using Aggregated Token Attention.
- Providing insights for practical deployment and e-commerce applications.

Introduction

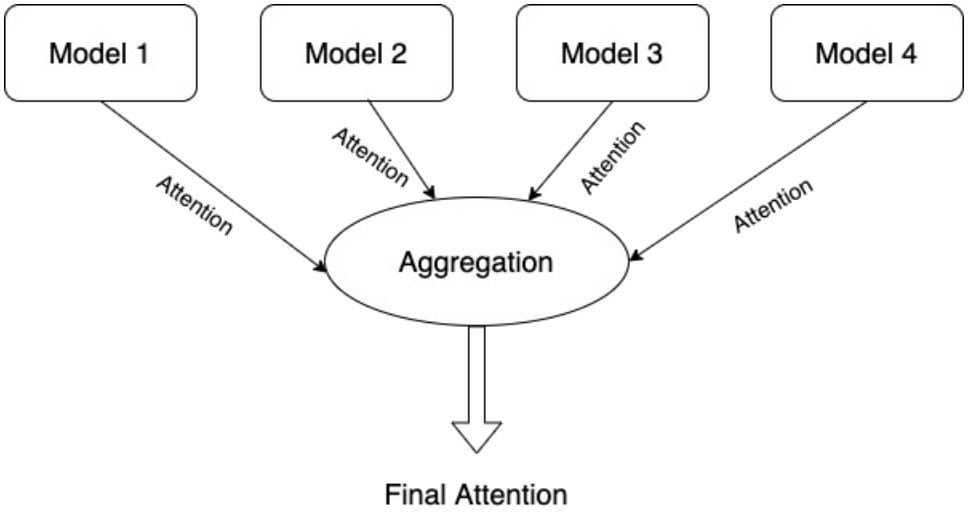
- With the rise of e-commerce, analyzing customer-generated content is critical for improving products and services.
- Robust sentiment classification
- Combining textual, user-specific, and rating data for enhanced performance.

Related Work

- Early sentiment analysis
    - SVM and Naive Bayes.
- BERT introduced superior performance by capturing token attention.
    - Devlin et al. [2019]
- Incorporating metadata, such as user profiles and ratings, further improves classification.

Methodology: Aggregated Token Attention Network (ATAN)

- Fine-tuning pre-trained BERT models
- Aggregating attention scores across models.
- Multi-model approach enhances robustness and accuracy through averaged token-level attention.



Methodology Details

- Fine-tuning: Models are trained on text reviews and user-specific ratings.
- Attention Aggregation: Token-level attention scores from multiple models are averaged.
- Sentiment Prediction: Class predictions are based on aggregated attention scores.

Experiment: Data Preprocessing

- Punctuation removal
- Unwanted word removal (e.g., URLs)
- Lowercase conversion
- Stop word removal
- Lemmatization to retain semantic meaning.

Experiment: Feature Engineering

- Down-sampling to address class imbalance
- Balanced dataset with 20,000 samples per class.
- The final dataset of 60,000 examples
- Split 80:20 for training and testing.

Evaluation Metrics

- Precision: Proportion of correctly predicted examples out of all positive predictions.
- Recall: Proportion of correctly predicted examples out of all actual instances.
- F1-Score: Harmonic mean of precision and recall for balanced evaluation.

Results: Model Comparisons

- BERT-Base-Uncased:
  - Accuracy 75%, F1-Score 0.76
- lxyuan/distilbert-base-multilingual-cased-sentiments-student:
  - Accuracy 71%, F1-Score 0.72
- nlptown/bert-base-multilingual-uncased-sentiment:
  - Accuracy 74%, F1-Score 0.75

Model	Accuracy	Precision (Macro Avg)	Recall (Macro Avg)	F1-Score (Macro Avg)	Support
BERT-Base-Uncased	0.75	0.76	0.75	0.76	12000
lxyuan/distilbert-base-multilingual-cased-sentiments-student	0.71	0.72	0.71	0.72	12000
nlptown/bert-base-multilingual-uncased-sentiment	0.74	0.75	0.74	0.75	12000
Proposed Model	**0.77**	**0.78**	**0.77**	**0.77**	**12000**

Results: Model Comparisons

```
ID_TO_LABEL = {0: 'Negative', 1: 'Neutral', 2: 'Positive'}
LABEL_TO_ID = {'Negative': 0, 'Neutral': 1, 'Positive': 2}
```

	precision	recall	f1-score	support
0	0.79	0.70	0.74	3958
1	0.63	0.73	0.68	4020
2	0.86	0.83	0.85	4022
accuracy			0.75	12000
macro avg	0.76	0.75	0.76	12000
weighted avg	0.76	0.75	0.76	12000

	precision	recall	f1-score	support
0	0.69	0.75	0.72	3958
1	0.62	0.64	0.63	4020
2	0.85	0.75	0.80	4022
accuracy			0.71	12000
macro avg	0.72	0.71	0.72	12000
weighted avg	0.72	0.71	0.72	12000

	precision	recall	f1-score	support
0	0.75	0.73	0.74	3958
1	0.64	0.68	0.66	4020
2	0.85	0.82	0.84	4022
accuracy			0.74	12000
macro avg	0.75	0.74	0.75	12000
weighted avg	0.75	0.74	0.75	12000

	precision	recall	f1-score	support
0	0.80	0.71	0.75	3958
1	0.65	0.75	0.69	4020
2	0.88	0.84	0.86	4022
accuracy			0.77	12000
macro avg	0.78	0.77	0.77	12000
weighted avg	0.78	0.77	0.77	12000

Challenges

- Computationally expensive
- Each model has specific tokenizer
  - Element-wise average?
    - Padding

Conclusion

- Study shows the significance of attention aggregation and preprocessing, feature engineering, and balancing datasets as well.
- The proposed model achieved superior performance and provides insights for deploying efficient sentiment analysis.
- Future work will explore multiple datasets and model interpretability.

References

- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. 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), pages 4171–4186, 2019.

Thank You