Supervisor :
Dr. Surangika Ranathunga

Dataset and Baseline for Automatic Student Feedback Analysis

H.M.M.V.B Herath
R.A.K.C Nilanga
M.K.H Maduwantha

PROBLEM

- STUDENT FEEDBACK CONSISTS OF TWO TYPES, QUALITATIVE FEEDBACK AND QUANTITATIVE FEEDBACK. QUALITATIVE FEEDBACK IS IMPORTANT TO GET A BETTER UNDERSTANDING OF THE STUDENTS' OPINION. MANUALLY ANALYZING THEM IS A VERY TEDIOUS TASK.

MOTIVATION

- MOST OF THE SENTIMENT ANALYSIS RESEARCH ON STUDENT FEEDBACK ANALYSIS HAS USED QUANTITATIVE FEEDBACK RATHER THAN QUALITATIVE FEEDBACK.
- NO PUBLIC STUDENT FEEDBACK DATASET HAS BEEN PUBLISHED YET.

OBJECTIVE

- CREATING A RELIABLE AND USEFUL STUDENT FEEDBACK CORPUS.
- ANALYSING FEEDBACK USING 3 SUBTASKS; TOWE, ALSA AND DLSA, BY USING SOTA TECHNIQUES FOR EACH TASK.
- RELEASE THE ANNOTATED DATASET TO THE PUBLIC.

METHODOLOGY DATASET

Dataset content

The dataset includes 3000 feedback collected from 2 sources;

- 1,379- from university student feedback
- 1,621- from ratemyproffesor.com website

**Problems in other student feedback annotation schemes [4,5]**

- Ambiguity exists between aspect categories because some aspects are closely related to each other.
- Only focus on lecturers' aspects but there are other aspects that are significant to the education domain.

Proposed solutions for the above problems

- Use a hierarchical taxonomy for aspects
- Introduction of new aspects such as "course structure", "subject materials", and "assessments"

Release of the dataset

- The annotated dataset is publicly released on Hugging Face, for the use of future research.
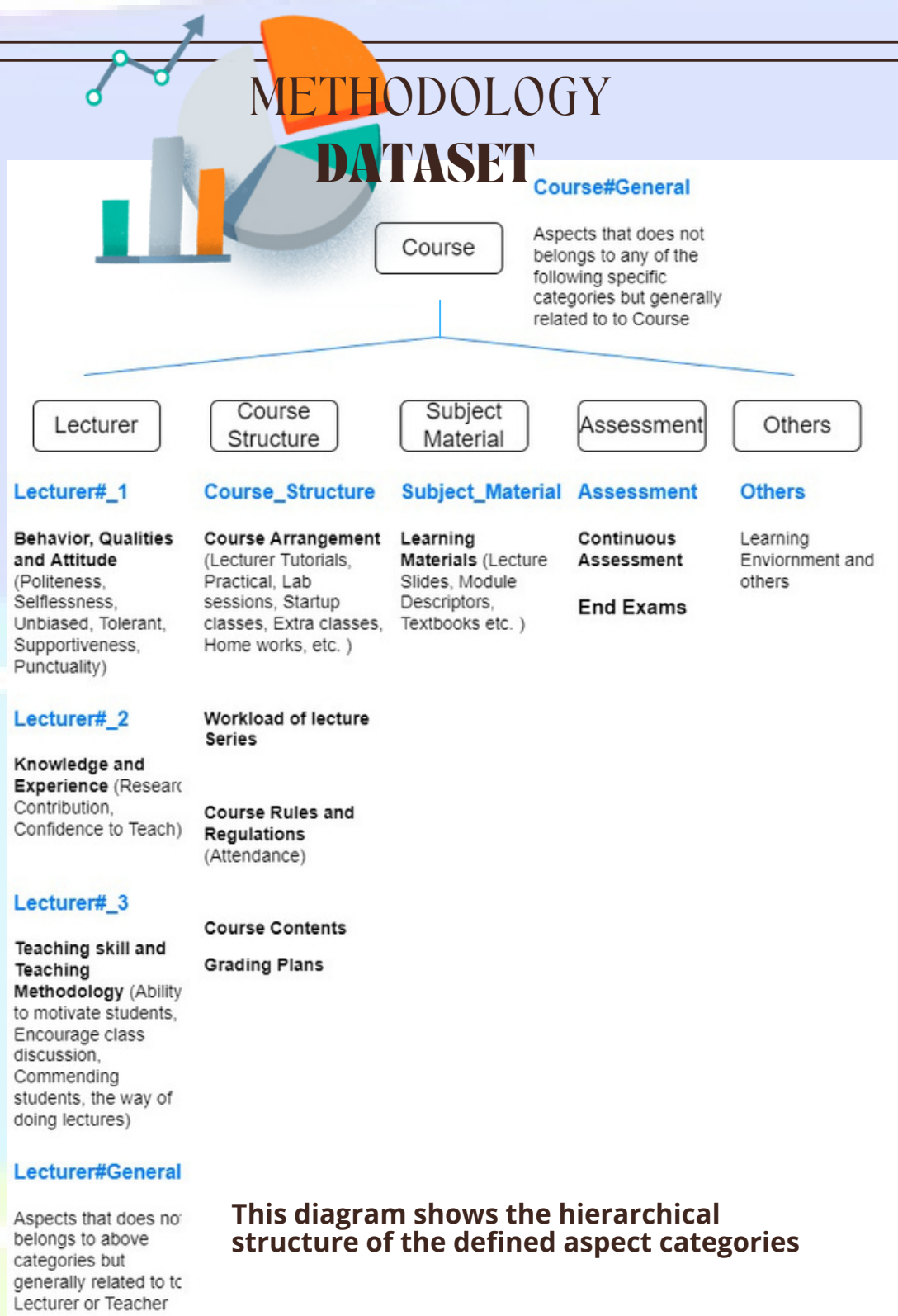
This diagram shows the hierarchical structure of the defined aspect categories

Sentiment polarities for ;
Aspect Level :
- Positive
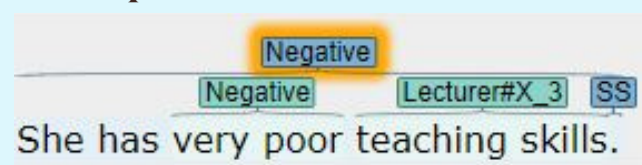- Negative
- Neutral=Sug
- Neutral=NSug

Document Level :
- Positive
- Negative
- Neutral

Dataset Distribution

Explicit Aspects: 81.93%
Implicit Aspects: 18.03%
Minimum Length: 1 word
Maximum Length: 493 words
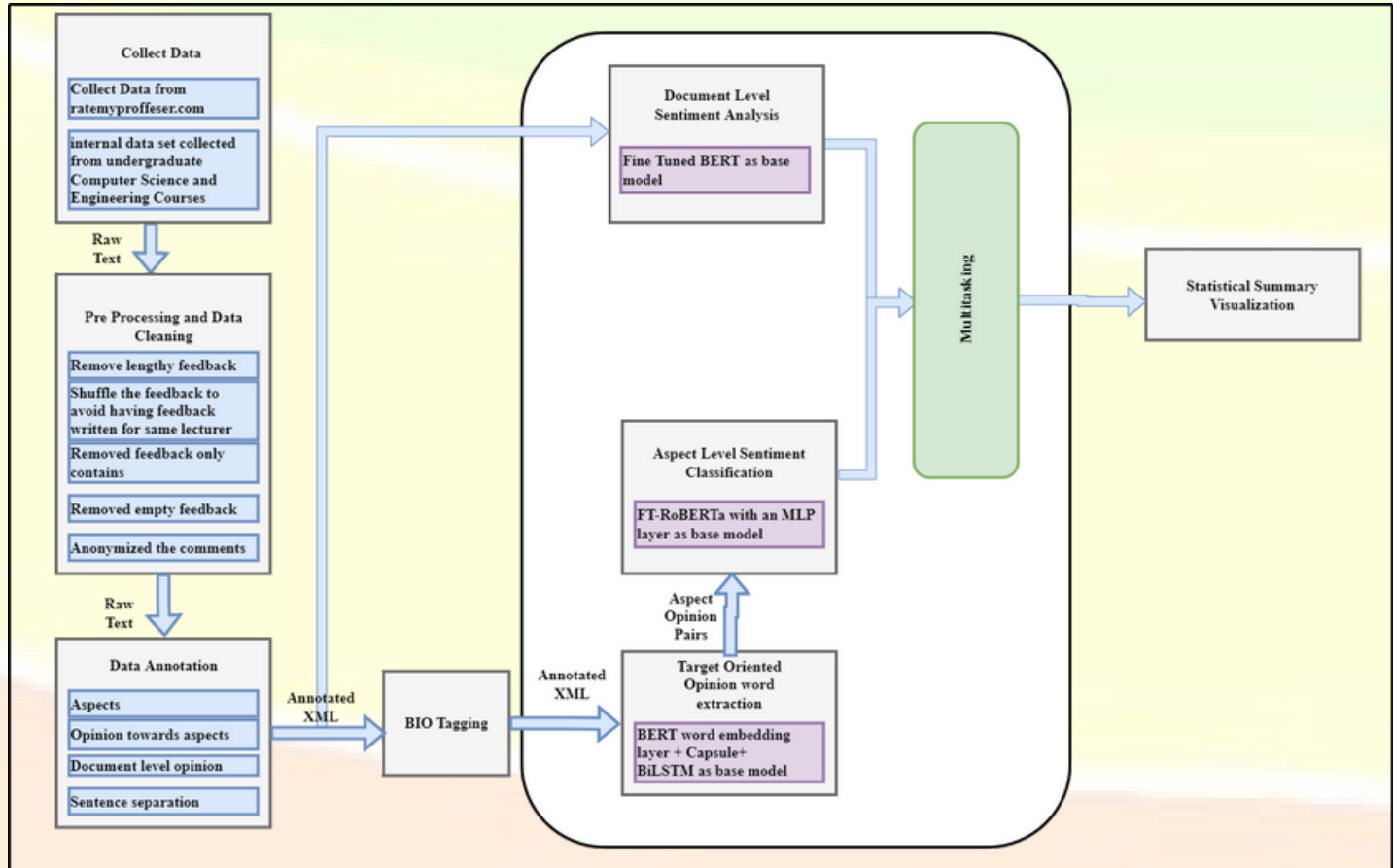Average Length: 26 words

Example -

She has very poor teaching skills.

ASPECT LEVEL
Opinion: very poor
Opinion Target: teaching skills
Aspect Category: Lecturer#3Explicit
Sentiment: Negative

DOCUMENT LEVEL
Sentiment: Negative

HIGH-LEVEL ARCHITECTURE OF OVERALL SYSTEM

TARGET - ORIENTED OPINION WORD EXTRACTION

The task of the Target Oriented Word Extraction (TOWE) model is to extract opinion words towards underlined aspect words [1]

ASPECT-LEVEL SENTIMENT ANALYSIS

The task of the ALSA [2] model is to predict the sentiment polarity expressed from the opinion words, towards the aspect words when the opinion words and aspect words are given.

DOCUMENT-LEVEL SENTIMENT ANALYSIS

The object of the DLSA [3] is to determine the overall opinion of the document.
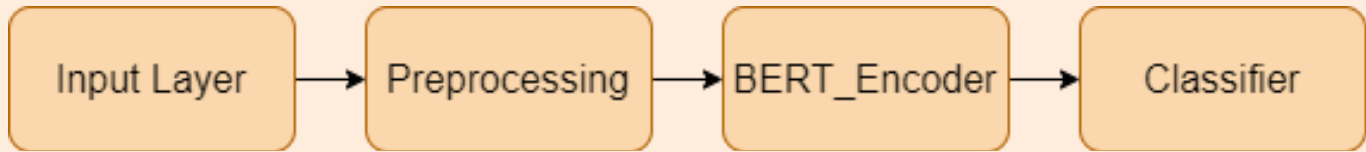
We train a BERT model to classify student feedback as positive, neutral, or negative, based on the overall polarity of the feedback.

We also fine-tune other BERT variants for this task.
- ALBERT
- RoBERTa
- XLNET

TESTING AND RESULTS

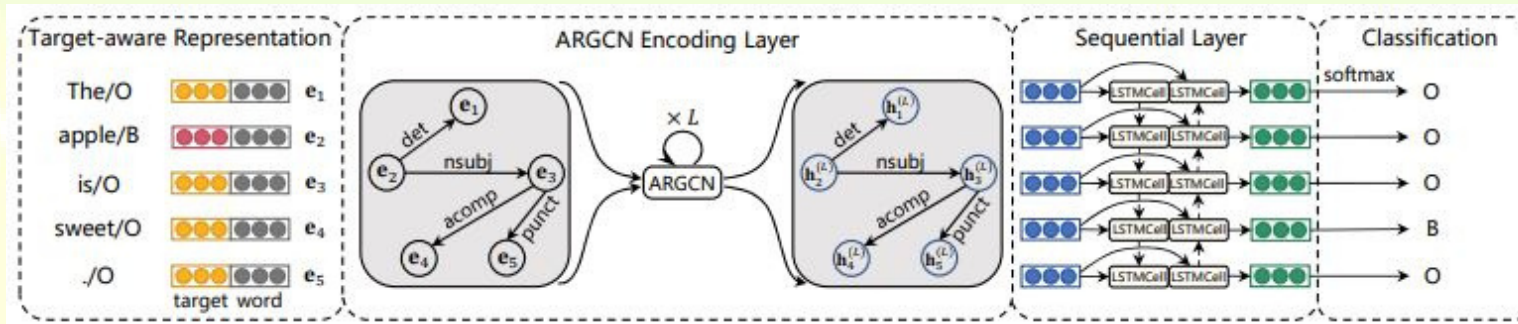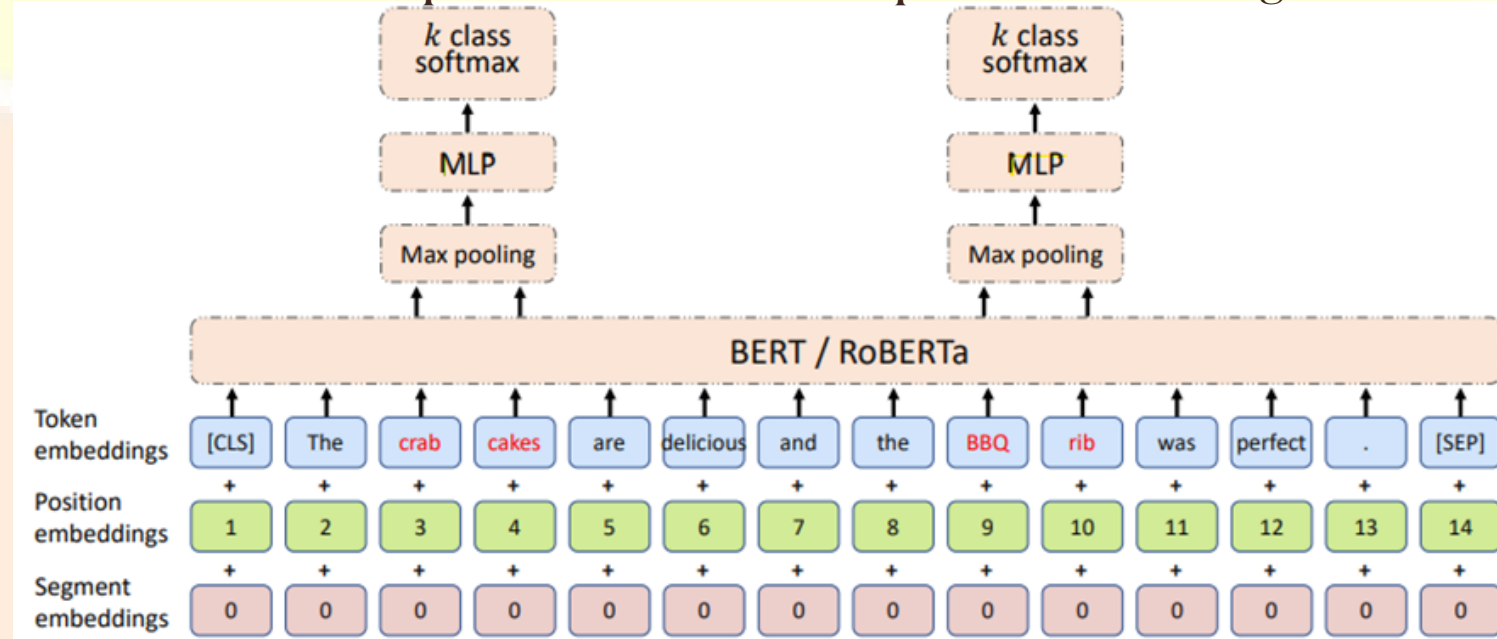| Task	Technique	Acc.	Prec.	Rec.	F1
	Baseline Models				
TOWE	Attention-based Relational Graph Convolutional Network	62%	65%	64%	64%
ALSA	RoBERTa	86%	78%	74%	75%
DLSA	BERT-base	86%	84%	83%	83%
	Single Task				
DLSA	ALBERT-base	85%	82%	82%	82%
	RoBERTa-base	86%	82%	80%	81%
	XLNET-base	86%	85%	82%	83%
	Tree Models				
ALSA	FT-BERT induced Tree + PWCN	78%	-	-	65%
	FT-BERT induced Tree + RGAT	82%	-	-	68%
	FT-RoBERTa induced Tree+ PWCN	79%	-	-	67%
	FT-RoBERTa induced Tree+ RGAT	83%	-	-	71%

CONCLUSION

- This paper dicuss about creating a student feedback corpus that can be used in NLP tasks such as;
  - Aspect Extraction,
  - Opinion Extraction ,
  - Aspect Opinion Pair Extraction and
  - Target Oriented Opinion Word Extraction etc.
- we use 3 baselines as TOWE, ALSA, and DLSA to ensure the reliability of dataset.
- For ALSA, a RoBERTa model was used with an MLP layer to induce trees using perturbed masking. A BERT model was used for the DLSA task and compared with the other BERT variants(XLNet, RoBERTa, ALBERT).
- An annotation scheme for the student feedback domain was created and the annotated dataset will be publicly released.

DATASET EVALUATION METRICS

Evaluation Metrics	Classification	Value
Fleiss Kappa	Document-Level	80%
Kripendorff's Kappa	Aspect	61%
	Opinion	66%

REFERENCE

[1] Jiang, J., Wang, A., & Aizawa, A. (2021, April). Attention-based Relational Graph Convolutional Network for Target-Oriented Opinion Words Extraction. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume (pp. 1986-1997).
[2] Dai, J., Yan, H., Sun, T., Liu, P. and Qiu, X., 2021. Does syntax matter? A strong baseline for Aspect-based Sentiment Analysis with RoBERTa. arXiv preprint arXiv:2104.04986.
[3] J. Devlin, M. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", arXiv.org, 2018. [Online]. Available: https://arxiv.org/abs/1810.04805.
[4] Sindhu, I., Muhammad Daudpota, S., Badar, K., Bakht-yar, M., Baber, J., and Nurunnabi, M. (2019).Aspect-based opinion mining on student's feedback for faculty teaching performance evaluation. IEEEAccess, 7:108729–108741.
[5] Chathuranga, J., Ediriweera, S., Hasantha, R., Munas-inghe, P., and Ranathunga, S. (2018). Annotating opinions and opinion targets in student course feed-back. In Proceedings of the Eleventh InternationalConference on Language Resources and Evaluation(LREC 2018)