

Dataset and Baseline for Automatic Student Feedback Analysis

Kushan Nilanga, Missaka Herath, Hashan Maduwantha, Surangika Ranathunga

Department of Computer Science and Engineering

University of Moratuwa, Katubedda 10400, Sri Lanka.

{kushanchamindu.17, mkhmaduwantha.17, missakaherath.17, surangika}@cse.mrt.ac.lk

Abstract

In this paper, we present a student feedback corpus that contains 3000 instances of feedback written by university students. This dataset has been annotated for aspect terms, opinion terms, polarities of the opinion terms towards targeted aspects, and document-level opinion polarities. We developed a hierarchical taxonomy for aspect categorisation, which covers many aspects of the teaching-learning process. We annotated both implicit and explicit aspects using this taxonomy. Annotation methodology, difficulties faced during the annotation, and the details of the aspect term categorization are discussed in detail. Using state-of-the-art techniques, we have built baseline models for the following tasks: Target oriented Opinion Extraction, Aspect Level Sentiment Analysis, and Document Level Sentiment Analysis. These models reported 64%, 75%, and 86% F1 scores (respectively) for the considered tasks. These results illustrate the reliability and usability of the corpus for different tasks related to sentiment analysis.

Keywords: Target-oriented Opinion Word Extraction, Aspect-level Sentiment Analysis, Document-level Sentiment Analysis, Pre-Trained Language Models (PLM), Student Feedback

1. Introduction

Student feedback is very useful when it comes to improving the teaching and learning process. Traditionally, student feedback was collected using forms that had to be manually read and summarized. But such manual analysis requires considerable time and effort. Nowadays online portals are available to collect feedback, but technology to automatically summarize/aggregate student feedback is not publicly available.

While document level sentiment analysis (DLSA) provides an overall idea of the sentiment of student feedback, it does not indicate the student’s opinion on different aspects of the the teaching/learning process. The solution is aspect level sentiment analysis (ALSA). ALSA may involve multiple sub-tasks: Aspect Extraction (AE), Opinion Extraction (OE), Aspect Opinion Pair Extraction (AOPE) and Target Oriented Opinion Word Extraction (TOWE).

Accordingly, student feedback can be analysed in the following manner:

1. Identifying and extracting all the aspects of a given feedback (AE)
2. Identifying and extracting all the opinions for a given aspect (TOWE)
3. Determining the sentimental polarity of the aspects (ALSA)
4. Determining the sentimental polarity of the whole feedback (DLSA)

In this paper, we only focus on the last three tasks (TOWE, ALSA, and DLSA, respectively).

In order to understand these tasks, consider the sentence, “Teaching skills are outstanding, and the exam was easy”,

- The task of the TOWE model is to extract opinion words towards aspect words (underlined) in the sentence. Here, it should extract *outstanding* and *easy* as the opinion words.
- The task of the ALSA model is to predict the sentiment polarity expressed by the opinion words, towards the aspect words when the opinion words and aspect words are given. Here we see a *positive* polarity towards the *Teaching skills* and *exam* aspects.
- The task of the DLSA model is to predict the overall sentiment of the feedback. For the given example, the model should predict a *positive* polarity.

Though there is a large amount of research conducted in the area of analyzing customer reviews (mainly using sentiment analysis techniques), there is very less amount of research conducted on analyzing student feedback. Some previous research in the student feedback domain has identified taxonomies of opinion targets (Sindhu et al., 2019; Chathuranga et al., 2018), but those are not comprehensive. We are not aware of any publicly available annotated datasets for this domain. We prepared a student feedback dataset annotated with a comprehensive set of opinion categories and opinion targets towards which the students have expressed their opinions. This corpus consists of 3000 student feedback, with 1800 feedback collected from <https://www.ratemyprofessors.com/> and 1200 from University of Moratuwa, Sri Lanka. This corpus has been publicly released and can be used for all the three tasks mentioned above.

To define aspect categories in the annotation schema, we referred to Sindhu et al. (2019) and Chathuranga et al. (2018). By further improving these annotation schema, we developed a new set of aspect categories that covers different areas of the teaching/learning process. We annotated both explicit and implicit aspect terms and corresponding opinions. Annotation of sentiment polarities has been divided into two parts: aspect level opinion and document level opinion.

To annotate document-level sentiment polarities, categories used in the SemEval 2016 shared task have been used (Positive, Negative, and Neutral) (Pontiki et al. (2016)). To annotate aspect-level sentiments, we introduced two new labels instead of “Neutral” as, “Neutral with Suggestions” (NeutralSug) and “Neutral without Suggestions” (NeutralNSug). This is because student feedback data contains a lot of suggestions and recommendations for fellow colleagues.

For Target Oriented Opinion Word Extraction, we chose (Jiang et al., 2021) as the baseline. For Aspect Level Sentiment Analysis (ALSA), we used Dai et al. (2021) as the baseline. For Document Level Sentiment Analysis (DLSA), we fine-tuned several pre-trained models that have shown promising results for related tasks - BERT-base (Xu et al., 2019), RoBERTa-base (Liu et al., 2019a), ALBERT-base (Lan et al., 2019) and XLNET-base (Yang et al., 2019).

In addition to the annotated dataset and code for reading, annotated data ¹ is publicly released, in the hope that it would help more research to focus on this relatively under-studied domain.

2. Related Work

2.1. Datasets and Taxonomies

Sindhu et al. (2019) developed an annotation scheme for student feedback analysis by conducting interviews with education domain experts. They introduced 12 aspect categories. However, this taxonomy has an issue of overlapping aspect categories. For example, “Behavior” aspect category can be overlapped with “Politeness” aspect category, since politeness is a type of behavior (Behavior aspect category \cap Polite aspect category $\neq \emptyset$). Despite the fact that Chathuranga et al. (2018) only used aspects related to lecturers, they reduced ambiguity between aspects by adopting a hierarchical representation of the aspects. Sivakumar and Reddy (2017) introduced 7 aspect categories including Teaching, Facilities, Sports, Fees etc. But when it comes to some aspects such as teaching, there can be fine-grained sub-aspects that need to be considered. However, such fine-grained aspects are not considered in their taxonomy. Lwin et al. (2020) also annotated a student feedback dataset, but did not develop a taxonomy for aspects.

¹https://huggingface.co/datasets/NLPC-UOM/Student_feedback_analysis_dataset

Van Nguyen et al. (2018) and Nguyen et al. (2018) both introduced student feedback corpora for the Vietnamese language, which is a low resource language. They also present the methods of building annotation guidelines that are used to ensure the annotation accuracy and the consistency of the corpus. The datasets are annotated for two different tasks: sentiment-based and topic-based classifications, with one label per sentence. None of this research has publicly shared their annotated dataset.

A noticeable amount of research has been done for other domains such as movies, restaurant and product purchasing (Wu et al., 2020), because there exist a significant amount of datasets with or without annotation in those domains such as SemEval 14/15/16 ¹, Amazon Review Data. ², Yelp ³ and IMDB ⁴ datasets.

2.2. Sentiment Analysis on Student Feedback

Lwin et al. (2020) used Naïve Bayes and Support Vector Machines classifiers for sentiment analysis on student feedback, while Sivakumar and Reddy (2017) carried out sentiment analysis based on aspects and classified sentences into the predefined set of aspect categories. They experimented with Decision Trees, Support Vector Machines and Naïve Bayes classifiers. Chathuranga et al. (2017) focused on extracting the opinion targets from student feedback using a Conditional Random Fields (CRF) classifier. Sindhu et al. (2019) used a two-layered LSTM model for aspect-based opinion mining and sentiment analysis. Misuraca et al. (2021) and Karunya et al. (2020) used an analytical strategy to automatically evaluate student feedback.

2.3. Aspect-level sentiment Analysis (ALSA)

According to Sindhu et al. (2019), theirs is the first attempt to use Deep Learning techniques for sentiment analysis of student feedback. A two-layered LSTM model for aspect extraction and sentiment classification was used in their method.

More recent work rely on solutions based on pre-trained language models (PLMs). A very recent work by Dai et al. (2021) used induced trees from a fine-tuned RoBERTa (Liu et al. (2019b)) model for ALSA. Their experiments showed that a RoBERTa-based model can achieve state-of-the-art performance in ALSA since it implicitly incorporates the syntactic information.

While Aspect term extraction and ALSA can be carried out as independent tasks, it is always possible to combine them into a multi-task setup. Some such solutions

¹<https://alt.qcri.org/semeval2016/task5/index.php?id=data-and-tools>

²<https://nijianmo.github.io/amazon/index.html>

³<https://www.yelp.com/dataset>

⁴<https://www.imdb.com/interfaces/>

have been given by He et al. (2019) and Chen et al. (2020).

2.4. Target-oriented opinion word extraction (TOWE)

TOWE task was initially proposed by Fan et al. (2019). TOWE refers to the sequence labeling task of extracting the opinion word towards the target word (aspect) from input sentences that contain n number of words and an opinion target (aspect). Fan et al. (2019) used Inward-LSTM (left context) and an Outward-LSTM (right context) model to capture both left and right context information and then use an IO-LSTM to concatenate both collected left and right context information. To the best of our knowledge, the state-of-art model for the TOWE task is Jiang et al. (2021), where they used an attention-based graph convolution network with BERT embeddings to achieve state-of-the-art performance.

2.5. Document-level sentiment Analysis (DLSA)

The objective of the DLSA task is to determine the overall opinion of the document. Document level sentiment analysis assumes that each document expresses opinions on a single entity. The BERT model (Xu et al. (2019)) and its variants (Lan et al. (2019), Liu et al. (2019a)), inspired by Transformers (Vaswani et al. (2017)), have thoroughly established themselves as the state-of-the-art methods in numerous Natural Language Processing tasks, including document-level sentiment analysis. Minaee et al. (2020a) extensively described Deep Learning based sentiment analysis models evaluated on the the Amazon Review ², Yelp ³ and IMDB data sets and showed that BERT and its variants achieve the best performance.

3. Dataset Creation

3.1. Data Collection

We collected data from two main resources;

1. ratemyprofessors.com⁶ - 20,001 feedback samples. In this portal, feedback has been grouped and sorted according to the name of the professor.
2. Student feedback collected from the students of University of Moratuwa - 1,379 feedback samples. In this dataset, feedback is grouped according to each course module and lecturer.

After pre-processing (see next sub-section), 3000 student feedback are randomly selected for annotation. The final dataset contains 1379 student feedback from University of Moratuwa and the rest of the feedback are from ratemyprofessors.com (1621 feedback).

²<https://nijianmo.github.io/amazon/index.html>

³<https://www.yelp.com/dataset>

⁶<https://www.ratemyprofessors.com/>

3.2. Data Pre-processing

As preprocessing steps, we randomized the data collected from ratemyprofessor.com to gather comments for random professors because overall sentiment for a particular professor can be unique. Then empty comments and comments that contain only symbols were removed from the dataset. Finally, to ensure privacy and confidentiality of data, we anonymized all the student feedback we collected from University of Moratuwa (data collected from ratemyprofessor.com were not anonymized as those are publicly available).

3.3. Data Annotation

There were three annotators and each annotator annotated 1000 student feedback. Before annotating data, the annotation scheme was created and was clearly explained to the annotators with detailed information. It helped annotators to annotate data easily. The dataset was annotated using the Inception tool Klie et al. (2018), in such a manner that the annotated dataset can be used for Aspect Opinion Pair Extraction, Opinion Target Extraction, Aspect Extraction and Opinion Extraction.

When annotating, we considered all implicit and explicit aspects because student feedback contains a considerable amount of aspects expressed in an implicit manner. The explicit aspect is one that appears in a sentence as a noun or noun phrase, whereas the implicit aspect is one that is implied in the sentence (Lal and Asnani, 2014). Table 1 contains statistics of all the tags that were used for annotation. Figure 1 describes the hierarchical structure of the defined aspect categories.

- Document Level Sentiment

For the document-level-sentiment analysis task, text is classified into Positive, Negative and Neutral polarities, considering the whole feedback. Following are some examples with their related class.

- Positive - The whole comment gives a positive feedback.
Eg: Madam taught well.
- Negative - The whole comment gives a negative feedback.
Eg: RUDE! Not helpful at all.
- Neutral - The whole comment does not give a positive or a negative feedback.
Eg: Madam, if possible, please include tutorials too.

- Aspect Level Sentiment

For Aspect Level sentiment classification, 4 opinion polarity categories have been considered in the annotation schema. Positive and Negative polarities were taken by following the annotation schema proposed by Chaturanga et al. (2018). Since feedback in the education domain consists

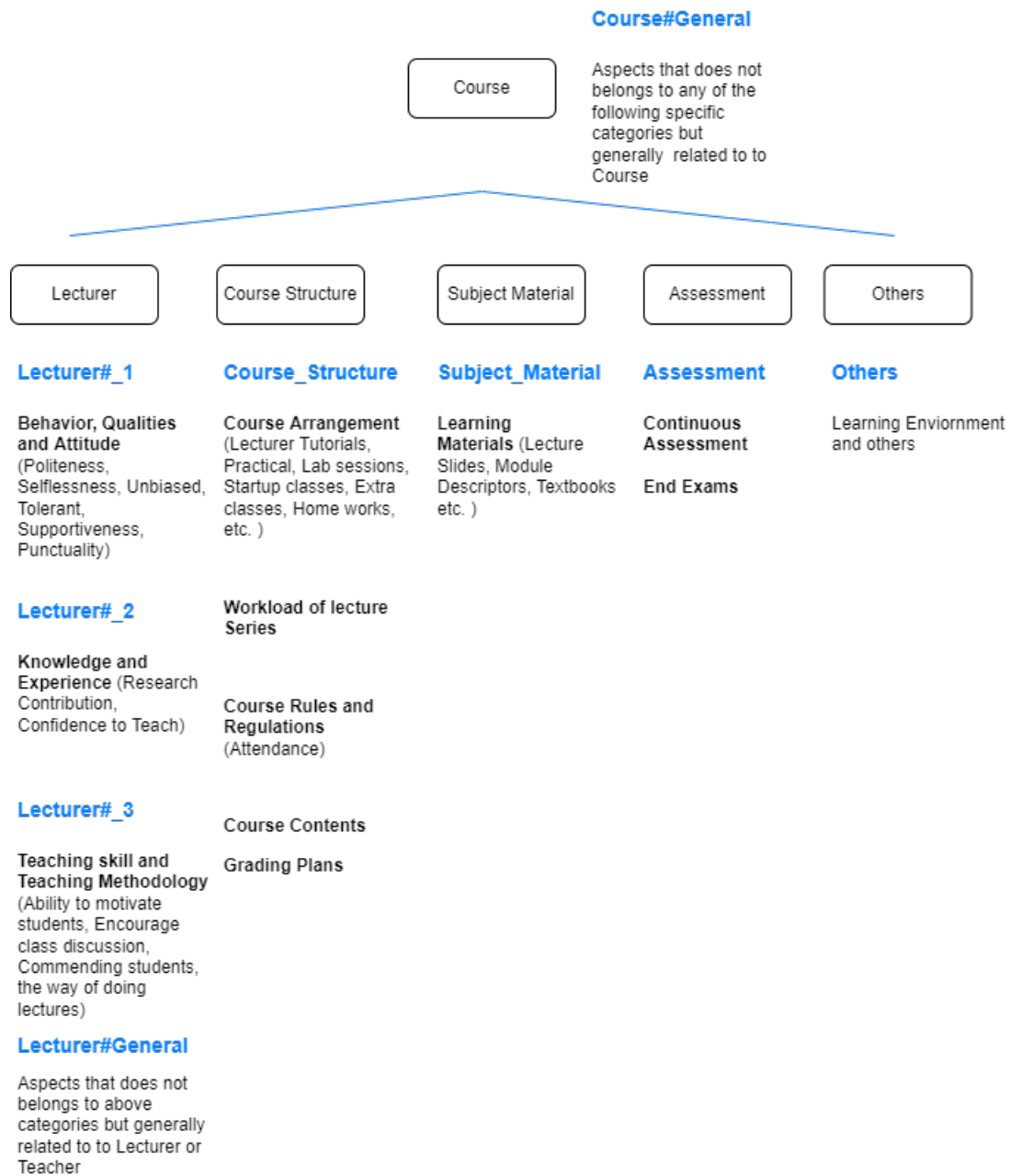


Figure 1: Hierarchically Structure of Aspect Categories

of a considerable amount of feedback that includes suggestions, 2 other opinion polarities were added to represent feedback with suggestions and without suggestions. The polarity of those two categories was considered neutral.

- Positive : Polarity of the opinion word, towards the aspect word is positive
 - * Eg: The course contents are good.
 - * Aspect word : course contents
 - * Aspect category : Course Structure
 - * Opinion word : good
 - * The opinion is positive towards “course contents”

- Negative : Polarity of the opinion word, towards the aspect word is negative
 - * Eg: The end exam is difficult.
 - * Aspect word : end exam
 - * Aspect category : Assesment
 - * Opinion word : difficult
 - * The opinion is negative towards “end exam”
- Neutral#Sug : Polarity of the opinion word towards the aspect word cannot be considered as either negative or positive. Also the opinion gives a suggestion.
 - * Eg: please give code examples with the

Classification	Tags	Label Count	percentage(%)
Document level opinion	Positive	1754	59.27
	Negative	561	18.95
	Neutral	644	21.76
Opinion	Positive	5152	59.82
	Negative	1832	21.27
	Neural with suggestion	624	7.24
	Neural without suggestion	1004	11.65
Aspect	Lecturer#1_Explicit	1448	17.40
	Lecturer#1_Implicit	66	0.79
	Lecturer#2_Explicit	97	1.17
	Lecturer#2_Implicit	42	0.50
	Lecturer#3_Explicit	1373	16.50
	Lecturer#3_Implicit	952	11.44
	Lecturer#General_Explicit	1353	16.26
	Lecturer#General_Implicit	139	1.67
	Course#General_Explicit	523	6.28
	Course#General_Implicit	167	2.01
	Course_Structure#Explicit	832	10.00
	Course_Structure#Implicit	54	0.65
	Subject_Material#Explicit	531	6.38
	Assessment#Explicit	532	6.39
	Assessment#Implicit	13	0.16
	Others#Explicit	130	1.56
	Others#Implicit	68	0.81
None	3	0.04	

Table 1: Dataset Statistics

Explicit aspect	81.93%
Implicit aspect	18.03%
Maximum Feedback Length	493
Minimum Feedback Length	1
Average Feedback Length	26

Table 2: Dataset Distribution

presentation.

- * Aspect word : code examples
- * Aspect category : Subject Material
- * Opinion word : give
- * The student is suggesting to give code examples
- Neutral#NSug : Polarity of the opinion word towards the aspect word cannot be considered as either negative or positive. The opinion does not give a suggestion.
 - * Eg: I learned many things all by myself.
 - * This sentence is neutral and the student does not suggest anything
- Aspects

The annotation scheme includes 13 aspect categories. To prevent ambiguity between aspect categories presented in Sindhu et al. (2019), we

re-grouped them in a hierarchical manner as in Chaturanga et al. (2018). Then following aspect categories were introduced, covering more fine-grained aspects in education domain.

- Lecturer#1 - Lecturer's behaviour and qualities
Eg: The lecturer was really kind and soft.
- Lecturer#2 - Lecturer's Knowledge and Experience such as research contribution and confidence
Eg: knowledgeable instructor.
- Lecturer#3 - Lecturer's teaching skill and teaching methodology such as the ability to motivate students, encouraging class discussion, commending students and the lecturing style
Eg: Sir, you have very good teaching skills. Thank you!
- Lecturer#General - Not belong to any of the above aspects of lecturer.
Eg: Terrible teacher.
- Course#General - Aspects that belong to the overall course or program
Eg: The course is really good.
- Course_Structure :

- * Course arrangement such as Lecture tutorials, lab sessions, practicals, startup class and extra classes.
Eg: The lecture series is structured very well.
 - * Workload of lecture series or program-related aspects
Eg: The workload is a bit too much
 - * Course rules and regulations related aspects such as attendance
Eg: Attendance is mandatory, there is a sign in sheet at the beginning.
 - * Course or program contents
Eg: Course had a very rich content of software development techniques.
 - * Lectures' grading criteria
Eg: He did NOT curve for our class but compensated (since the class average is an F) by giving us free points.
- Subject_Material - Learning material (lecture slides, module descriptors, etc.)
Eg: Lecture slides were easy to understand.
 - Assessment :
 - * Aspects related to the final exam or assessment
Eg: It's better if we could do more questions before the final exam.
 - * Aspects related to the continuous assessments such as mid exams, quizzes and assignments.
Eg: Quizzes were given every week so reading was a must and so was attendance.
 - Others - Aspects that do not belong to any of the above categories. Use as wildcard category
Eg: Most chairs were broken and loudspeakers are not clear

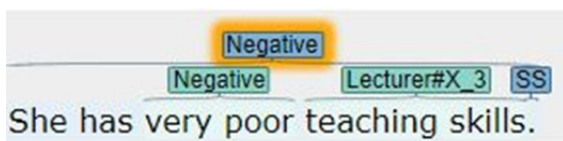


Figure 2: An annotated example

Figure 2 shows an example of an annotated feedback. Following are the details that are annotated in this feedback.

Aspect Level (there is only one aspect in this example);

- Opinion : very poor
- Opinion Target : teaching skills
- Aspect Category : Lecturer#3_Explicit

- Sentiment : Negative

Document Level;

- Sentiment : Negative

SS - Here SS tag is used to identify the sentence separation.

If we take the example “Good teacher, **Got the point very quickly..**”, the bold section tells a positive sentiment towards the lecturers knowledge. So we can extract the implicit aspect Lecturer#2_implicit that has a positive polarity in this comment.

3.4. Annotation Evaluation

The annotation process was carried out by three annotators. We calculated the inter-annotator agreement using Fleiss Kappa for single-label tags and Krippendorff’s alpha for multi-label tags. Table 3 displays the calculated evaluation matrices, and all of the values in the table are greater than 0.6, indicating that there is significant inter-annotator agreement among annotators.

Evaluation metrics	Classification	Value
Fleiss Kappa	Document level opinion	0.8008
	Sentence Separation	0.909
Krippendorff’s Alpha	Aspect	0.6132
	Opinion	0.6579

Table 3: Inter-Annotator Agreement Evaluation

4. Baselines

The baseline models we selected for each task are from the very recent research. They are all based on pre-trained language models.

4.1. Target-oriented Opinion Word Extraction (TOWE)

Attention-based Relational Graph Convolutional Network proposed by Jiang et al. (2021) was selected as the baseline for the Target-oriented Opinion Word Extraction task. This model uses BERT as a word embedding layer and generates a synthetic dependency graph based on Spacy dependency parser. Then an L-layer Attention-Based Graph Convolution network (A-GCN) is used for encoding synthetic and semantic information. The baseline also uses a Bi-LSTM (bi-directional Long Short Term Memory) model to process the sequential information as the TOWE is a sequential-labeling task. They have experimentally proven that three layers of A-GCN work much better for the considered domain.

4.2. Aspect-level Sentiment analysis

Dai et al. (2021) compared the induced trees from pre-trained Language Models (PLMs), fine-tuned

PLMs (FT-PLM), and the dependency parsing trees on several popular models for the ALSA task. For this, the researchers have used three representative tree-based ALSA models: Aspect-specific Graph Convolutional Networks (ASGCN), Proximity-Weighted Convolution Network (PWCN) and Relational Graph Attention Network (RGAT).

Perturbed masking techniques are used to induce trees from PLMs and FT-PLMs. We compare PWCN and RGAT structures induced from FT-BERT and FT-RoBERTa models that were fine-tuned with the student feedback dataset.

4.3. Document-level Sentiment Analysis

BERT and its variants have shown state-of-the-art performances in text classification benchmarks like IMDB movie reviews, Yelp, and Amazon datasets (Minaee et al., 2020b). In this study, we fine-tune the following models for the task of document-level sentiment analysis for student feedback: BERT-base (cased), XLNet-base, ALBERT-base and RoBERTa-base.

5. Results

Table 4 shows the best results obtained for Target-oriented Opinion Word Extraction, Aspect-level Sentiment analysis and Document-level Sentiment Analysis tasks for our student feedback dataset. According to the obtained results for aspect-level sentiment analysis mentioned in Table 5, FT-RoBERTa induced trees outperform the trees induced by other models. Table 6 shows the obtained results from the pre-trained BERT variant models for our dataset on document-level sentiment analysis task.

Baseline model	Acc.	Prec.	Rec.	F1
Target Oriented Opinion Extraction	0.62	0.65	0.64	0.64
Aspect-level Sentiment Analysis	0.86	0.78	0.74	0.75
Document-level Sentiment Analysis	0.86	0.84	0.83	0.83

Table 4: Results of the baseline models for each task

Pretrained model	Tree structure	Acc.	F1
FT-BERT Induced	PWCN	0.78	0.65
	RGAT	0.82	0.68
FT-RoBERTa Induced	PWCN	0.79	0.67
	RGAT	0.83	0.71

Table 5: The performance of tree-based ALSA models incorporating different tree structures on the student feedback dataset. FT-BERT, and FT-RoBERTa Induced Tree refer to tree structures induced from corresponding PLMs.

Pretrained model	Acc.	Prec.	Rec.	F1
BERT-base	0.86	0.84	0.83	0.83
ALBERT-base	0.85	0.82	0.82	0.82
RoBERTa-base	0.86	0.84	0.82	0.83
XLNet-base	0.86	0.85	0.82	0.83

Table 6: Results for the DLSA task

6. Conclusion

We presented a student feedback dataset annotated for a predefined set of aspect categories, opinion words and opinion word polarities and document level opinion polarities. We trained target-oriented opinion word extraction using an attention-based graph convolution model, aspect level sentiment analysis using a FT-RoBERTa model with an MLP layer, and document level sentiment analysis using a BERT based model. All these are state-of-the-art techniques used for the respective tasks, thus can be considered as strong baselines. One of the main limitations of the proposed system is that TOWE model does not identify aspects, and those should be annotated explicitly. As future work, the dataset should be annotated for aspects. We also intend to implement a multi-task model that is capable of performing all the three tasks mentioned above.

7. Acknowledgement

This research was partially funded by a Senate Research Committee (SRC) grant of University of Moratuwa, Sri Lanka.

8. Bibliographical References

- Chaturanga, J., Ediriweera, S., Munasinghe, P., Hasantha, R., and Ranathunga, S. (2017). Opinion target extraction for student course feedback. In *Proceedings of the 29th Conference on Computational Linguistics and Speech Processing (ROCLING 2017)*, pages 295–307.
- Chaturanga, J., Ediriweera, S., Hasantha, R., Munasinghe, P., and Ranathunga, S. (2018). Annotating opinions and opinion targets in student course feedback. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.
- Chen, S., Liu, J., Wang, Y., Zhang, W., and Chi, Z. (2020). Synchronous double-channel recurrent network for aspect-opinion pair extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6515–6524.
- 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*.
- Fan, Z., Wu, Z., Dai, X., Huang, S., and Chen, J. (2019). Target-oriented opinion words extraction with target-fused neural sequence labeling. 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 2509–2518.
- He, R., Lee, W. S., Ng, H. T., and Dahlmeier, D. (2019). An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. *arXiv preprint arXiv:1906.06906*.
- Jiang, J., Wang, A., and Aizawa, A. (2021). 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*, pages 1986–1997.
- Karunya, K., Aarthy, S., Karthika, R., and Deborah, L. J. (2020). Analysis of student feedback and recommendation to tutors. In *2020 International Conference on Communication and Signal Processing (ICCSP)*, pages 1579–1583. IEEE.
- Klie, J.-C., Bugert, M., Boullosa, B., de Castilho, R. E., and Gurevych, I. (2018). The inception platform: Machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9. Association for Computational Linguistics, June. Event Title: The 27th International Conference on Computational Linguistics (COLING 2018).
- Lal, M. and Asnani, K. (2014). Implicit aspect identification techniques for mining opinions: a survey. *International Journal of Computer Applications*, 98(4).
- Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., and Soricut, R. (2019). ALBERT: A lite BERT for self-supervised learning of language representations. *CoRR*, abs/1909.11942.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019a). Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019b). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Lwin, H. H., Oo, S., Ye, K. Z., Lin, K. K., Aung, W. P., and Ko, P. P. (2020). Feedback analysis in outcome base education using machine learning. In *2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pages 767–770. IEEE.
- Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., and Gao, J. (2020a). Deep learning based text classification: A comprehensive review. *CoRR*, abs/2004.03705.
- Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., and Gao, J. (2020b). Deep learning based text classification: A comprehensive review. *CoRR*, abs/2004.03705.
- Misuraca, M., Scepi, G., and Spano, M. (2021). Using opinion mining as an educational analytic: An integrated strategy for the analysis of students’ feedback. *Studies in Educational Evaluation*, 68:100979.
- Nguyen, P. X., Hong, T. T., Van Nguyen, K., and Nguyen, N. L.-T. (2018). Deep learning versus traditional classifiers on vietnamese students’ feedback corpus. In *2018 5th NAFOSTED conference on information and computer science (NICS)*, pages 75–80. IEEE.
- Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N., Jiménez-Zafra, S. M., and Eryigit, G. (2016). SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30, San Diego, California, June. Association for Computational Linguistics.
- Sindhu, I., Muhammad Daudpota, S., Badar, K., Bakhtyar, M., Baber, J., and Nurunnabi, M. (2019). Aspect-based opinion mining on student’s feedback for faculty teaching performance evaluation. *IEEE Access*, 7:108729–108741.
- Sivakumar, M. and Reddy, U. S. (2017). Aspect based sentiment analysis of students opinion using machine learning techniques. In *2017 International Conference on Inventive Computing and Informatics (ICICI)*, pages 726–731. IEEE.
- Van Nguyen, K., Nguyen, V. D., Nguyen, P. X., Truong, T. T., and Nguyen, N. L.-T. (2018). Uitvsfc: Vietnamese students’ feedback corpus for sentiment analysis. In *2018 10th international conference on knowledge and systems engineering (KSE)*, pages 19–24. IEEE.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. *CoRR*, abs/1706.03762.
- Wu, Z., Zhao, F., Dai, X.-Y., Huang, S., and Chen, J. (2020). Latent opinions transfer network for target-oriented opinion words extraction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 9298–9305.
- Xu, H., Liu, B., Shu, L., and Yu, P. S. (2019). Bert post-training for review reading comprehension and aspect-based sentiment analysis. *arXiv preprint arXiv:1904.02232*.
- Yang, Z., Dai, Z., Yang, Y., Carbonell, J. G., Salakhutdinov, R., and Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. *CoRR*, abs/1906.08237.