110-2 NLP Final Project

Tasks - Aspect-Based Sentiment Analysis

- Aspect-Based Sentiment Analysis aims to predict sentiment polarities on a set of predefined aspect categories mentioned in the input text.
- This final project consists of two tasks: Aspect Category Detection and Aspect Category Sentiment Classification.
  - Aspect Category Detection (Task1): Given an input text and a predefined set of aspect categories, identify the aspect categories discussed in the text.
  - Aspect Category Sentiment Classification (Task2): Given an input text and the aspect categories presented in the text, determine the polarity (positive, negative and neutral) of each aspect category discussed in the text.
- For example, given a set of aspect categories {Food, Service, Price, Ambience and Anecdotes/Miscellaneous}, “Although the fish is delicious, the waiter is horrible!” mentions Food and Service, the sentiment about the Food is positive while the opinion over the Service is negative.

Dataset

- The dataset is a slightly modified version of ASAP: A Chinese Review Dataset Towards Aspect Category Sentiment Analysis and Rating Prediction
- A large-scale Chinese restaurant review dataset including over 40k reviews from a leading online-to-offline (O2O) e-commerce platform in China. Besides a 5-star scale rating (not used in this final project), each review is manually annotated according to its sentiment polarities towards 18 pre-defined aspect categories.

Dataset

- Aspect Categories

Table 1: The full list of 18 aspect categories and definitions.

Aspect category	Definition	Aspect category	Definition
Food#Taste (口味)	Food taste	*Location#Easy_to_find* (是否容易寻找)	Whether the restaurant is easy to find
Food#Appearance (外观)	Food appearance	*Service#Queue* (排队时间)	Whether the queue time is acceptable
Food#Portion (分量)	Food portion	*Service#Hospitality* (服务人员态度)	Waiters/waitresses' attitude/hospitality
Food#Recommend (推荐程度)	Whether the food is worth being recommended	*Service#Parking* (停车方便)	Parking convenience
Price#Level (价格水平)	Price level	*Service#Timely* (点菜/上菜速度)	Order/Serving time
Price#Cost_effective (性价比)	Whether the restaurant is cost-effective	*Ambience#Decoration* (装修)	Decoration level
Price#Discount (折扣力度)	Discount strength	*Ambience#Noise* (嘈杂情况)	Whether the restaurant is noisy
Location#Downtown (位于商圈附近)	Whether the restaurant is located near downtown	*Ambience#Space* (就餐空间)	Dining Space and Seat Size
Location#Transportation (交通方便)	Convenient public transportation to the restaurant	*Ambience#Sanitary* (卫生情况)	Sanitary condition

Dataset Format

- The dataset has the following columns:
  - id: id for each review
  - review: text of the review (in simplified Chinese)
  - each of the 18 aspect categories:
    - -2: the aspect is not mentioned in the text
    - -1: negative
    - 0: neutral
    - 1: positive
- The dataset is uploaded to NTU COOL

Dataset Format

Table 3: A review example in ASAP, with overall star rating and aspect category sentiment polarity annotations.

Review	Rating	Aspect Category	Label	Aspect Category	Label
With convenient traffic, the restaurant holds a high-end decoration, but quite noisy because a wedding ceremony was being held in the main hall. Impressed by its delicate decoration and grand appearance though, we had to wait for a while at the weekend time. However, considering its high price level, the taste is unexpected. We ordered the Kung Pao Prawn, the taste was acceptable and the serving size is enough, but the shrimp is not fresh. In terms of service, you could not expect too much due to the massive customers there. By the way, the free-served fruit cup was nice. Generally speaking, it was a typical wedding banquet restaurant rather than a comfortable place to date with friends. 交通还挺方便的，环境看起来很高大上的样子，但是因为主厅在举办婚礼非常混乱，特别吵感觉，但是装修的还不错，感觉很精致的装修，门面很气派，周末去的时候还需要等位。味道的话我觉得还可以但是跟价格比起来就很一般了，性价比挺低的，为了去吃宫保虾球的，但是我觉得也就那样吧虾不是特别新鲜，不过虾球很大，味道还行。服务的话由于人很多所以也顾不过来上菜的速度不快，但是有送水果杯还挺好吃的。总之就是典型的婚宴餐厅不是适合普通朋友吃饭的地方了。	3-Star	*Location#Transportation* (交通方便)	1	*Price#Discount* (折扣力度)	-
		Location#Downtown (位于商圈附近)	-	*Ambience#Decoration* (装修)	1
		Location#Easy_to_find (是否容易寻找)	-	*Ambience#Noise* (嘈杂情况)	−1
		Service#Queue (排队时间)	-	*Ambience#Space* (就餐空间)	1
		Service#Hospitality (服务人员态度)	-	*Ambience#Sanitary* (卫生情况)	-
		Service#Parking (停车方便)	-	*Food#Portion* (分量)	1
		Service#Timely (点菜/上菜速度)	−1	*Food#Taste* (口味)	1
		Price#Level (价格水平)	0	*Food#Appearance* (外观)	-
		Price#Cost_effective (性价比)	−1	*Food#Recommend* (推荐程度)	-

Evaluation Metrics

- Given the dataset D, the ground-truth aspect categories of the target t ∈ D is denoted as Y_t, while the predicted aspect categories denoted as $\hat{Y_t}$
- Aspect Category Detection - Macro F1

$$\text{Macro-F1} = 2\frac{\text{Ma-P}\times\text{Ma-R}}{\text{Ma-P}+\text{Ma-R}}, \text{ which is based on Macro-Precision (Ma-P) and Micro-Recall (Ma-R) with Ma-P} = \frac{1}{|D|}\sum_{t\in D}\frac{|Y_t\cap\hat{Y}_t|}{\hat{Y}_t}, \text{ and Ma-R} = \frac{1}{|D|}\sum_{t\in D}\frac{|Y_t\cap\hat{Y}_t|}{Y_t}.$$

- Aspect Category Sentiment Classification - Accuracy
  - The number of correctly predicted aspect category polarity divided by the total number of aspect category polarity label
  - Aspect category not mentioned in the text is ignored in this phase

Grading

- Baseline: 30%
  - Your private scores have to pass the baseline in both tasks.
  - Baseline1 20% (LSTM-based model will be enough)
    - Task1-Macro F1: 0.75
    - Task2-Accuracy: 0.77
  - Baseline2 10% (BERT-based model will be enough)
    - Task1-Macro F1: 0.84
    - Task2-Accuracy: 0.84
- Private Leaderboard: 25% (12.5% for each task)
  - Rank 1-3 get 12.5, Rank 4-6 get 11, Rank 7-9 get 9.5, Rank 10-12 get 8, Rank 13-15 get 6.5, Rank 16-17 get 5
- Report and Code: 45%

Leaderboard Submission

- Kaggle sites:
  - Task1: https://www.kaggle.com/t/00cf1debdbac4028b7ce43a4c4285f80
  - Task2: https://www.kaggle.com/t/2551809e008b439fa2014f6ea5085ed2
- Participate as a team, your team name should be: team{YOUR_TEAM_NUMBER}
- No personal submissions
- Even if you predict 0 for an id-#aspect pair in task 1, you still have to submit a predicted sentiment in task2, we will ignore the pair if the aspect is not mentioned in the text according to the gold label.

Submission Format

- Task1
  - id-#aspect: review id and aspect pairs, for example, 1000-5 stands for the fifth aspect (Service#Hospitality) of the review with id 1000.
  - predicted: if you predict that the aspect is mentioned in the review, you should output the corresponding review id in your output file, otherwise you should output None. For example, if you predict that the fifth aspect (Service#Hospitality) is mentioned in the review with id 1000, you should output 1000, otherwise you should output None.
- Task2
  - id-#aspect: review id and aspect pairs, for example, 1000-5 stands for the fifth aspect (Service#Hospitality) of the review with id 1000.
  - sentiment: -1 for negative, 0 for neutral and 1 for positive

Report and Code

- You can organize your reports in any format but the report should include the following sections:
  - Method: preprocess, model architecture, pretrained model, etc.
  - Experiment: environment setting, hyperparameters, result, etc.
  - Workload Distribution
- Code
  - Provide a README with clear instructions of how to train or test your model
  - Don’t upload your model and data files
- Upload your report and code with the format: team{YOUR_TEAM_NAME}.zip
- One submission for each team

Rules

- No late submissions
- No cheating
- Use the dataset we provide, you can not use the original dataset in any way.
- You can use any other dataset and publicly-released model.

Timeline

- 5/5: Anouncement
- 6/10 23:59pm: Leaderboard Deadline
- 6/12 23:59pm: Report and Code Deadline