## A. Data Description

- 1. The Aspect\_terms folder includes files containing standard aspect terms, and the same or approximately standard aspect words on ontology are as follows:
  - ant aspect.xls: safety aspect.
  - doc aspect.xls: engine aspect.
  - gib aspect.xls: price aspect.
  - ngt aspect.xls: exterior aspect.
  - not aspect.xls: interior aspect.
  - tot\_aspect.xls: overall aspect.
  - vah aspect.xls: performance aspect.

Figure 1 shows the structure of these files.

1	Α	В	С				
1	index	aspect_item	similar_item				
2	1	an toàn	abs				
3	2	an toàn	an toàn				
4	3 an toàn		bám đường				

Figure 1. Structure of files containing aspect terms.

The structure of the files is as follows:

- *index* column: identification of pattern.
- aspect item column: the standard aspect terms.
- similar item column: the same or approximately standard aspect words.
- 2. The word2vec\_aspect folder includes the word2vec files. These files use for aspect detection tasks.
  - ant word size 300.model: safety aspect.
  - doc word size 300.model: engine aspect.
  - gib word size 300.model: price aspect
  - ngt word size 300.model: exterior aspect
  - not word size 300.model: interior aspect.
  - tot word size 300.model: overall aspect.
  - vah word size 300.model: performance aspect.
- 3. The Single\_Aspect\_Standardized\_Sentence folder includes the files containing data used in training and testing that are standardized. Each sentence has only one aspect.
  - The data files for the engine aspect in the two experiments are as follows:
  - Engine atempt 1.xls: The training data is the part from row 2 to row 34142.
  - Engine atempt 2.xls: The training data is the part from row 2 to row 34132.
  - The data files for the safety aspect in the two experiments are as follows:
  - Safety\_atempt\_1.xls: The training data is the part from row 2 to row 34165.
  - Safety atempt 2.xls: The training data is the part from row 2 to row 34160

- The data files for the price aspect in the two experiments are as follows:
- Price atempt 1.xls: The training data is the part from row 2 to row 34061.
- Price atempt 2.xls: The training data is the part from row 2 to row 34061.
- The data files for the exterior aspect in the two experiments are as follows:
- Exterior atempt 1.xls: The training data is the part from row 2 to row 34015.
- Exterior\_atempt\_2.xls: The training data is the part from row 2 to row 34015.
- The data files for the interior aspect in the two experiments are as follows:
- Interior atempt 1.xls: The training data is the part from row 2 to row 34161.
- Interior atempt 2.xls: The training data is the part from row 2 to row 34142.
- The data files for the overall aspect in the two experiments are as follows:
- Overall atempt 1.xls: The training data is the part from row 2 to row 33701.
- Overall atempt 2.xls: The training data is the part from row 2 to row 33701.
- The data files for the performance aspect in the two experiments are as follows:
- Performance atemp 1.xls: The training data is the part from row 2 to row 34074.
- Performance atemp 2.xls: The training data is the part from row 2 to row 34086.

Figure 2 shows the structure of these files.

1	Α	В	C
1	id	text	polarity
2	200000	cái_giá an_toàn	other
3	200001	1 anh xe_ôm chia_sẻ hài_hước	other
4	200002	1 chiếc minivan đủ đảm_bảo an_toàn tính_năng kiểm_soát xe hiện_đại	other

Figure 2. Structure of files containing training and testing data for aspects.

The structure of the files is as follows:

- *id* column: identification of pattern.
- *text* column: the content of sentences.
- polarity column: sentiment labels.
- 4. The Single\_Aspect\_Raw\_Sentence folder includes the files containing raw data used in training and testing. Each sentence has only one aspect.
  - The data files for the engine aspect in the two experiments are as follows:
  - Engine atempt 1 raw.xls: The training data is the part from row 2 to row 8593.
  - Engine\_atempt\_2\_raw.xls: The training data is the part from row 2 to row 8581.
  - The data files for the safety aspect in the two experiments are as follows:
  - Safety atempt 1 raw.xls: The training data is the part from row 2 to row 8613
  - Safety atempt 2 raw.xls: The training data is the part from row 2 to row 8610.
  - The data files for the price aspect in the two experiments are as follows:
  - Price atempt 1 raw.xls: The training data is the part from row 2 to row 8413.
  - Price atempt 2 raw.xls: The training data is the part from row 2 to row 8511.
  - The data files for the exterior aspect in the two experiments are as follows:
  - Exterior atempt 1 raw.xls: The training data is the part from row 2 to row 8465.
  - Exterior atempt 2 raw.xls: The training data is the part from row 2 to row 8465.
  - The data files for the interior aspect in the two experiments are as follows:

- Interior atempt 1 raw.xls: The training data is the part from row 2 to row 8611.
- Interior atempt 2 raw.xls: The training data is the part from row 2 to row 8613.
- The data files for the overall aspect in the two experiments are as follows:
- Overal atempt 1 raw.xls: The training data is the part from row 2 to row 8152.
- Overal atempt 2 raw.xls: The training data is the part from row 2 to row 8151.
- The data files for the overall aspect in the two experiments are as follows
- Performance atemp 1 raw.xls: The training data is the part from row 2 to row 8593.
- Performance atemp 2 raw.xls: The training data is the part from row 2 to row 8608.

These files have the same structure as shown in Figure 2.

5. The Data\_for\_sentiment\_analysis folder contains the sentiment\_data\_train\_3C\_doc.xls file for the sentiment analysis task. The training data is the part from row 2 to row 37269. Figure 3 shows the structure of this file.

4	Α	В	С
1	id	text	polarity
2	1000	1 chiếc minivan đủ sang_trọng đảm_bảo an_toàn tiện_dụng	pos
3	1001	1 chiếc xe được bán chính hãng sinh nó đã vậy không gọi xe độ cả	neu
4	1002	\$11m để làm chuyện khác hơn	neu
5	1003	1 chiếc xe hoàn_hảo	pos
6	1004	1 dàn lạnh người ngồi trước cóng người ngồi sau cuối sẽ mát	neg
7	1005	1 kiệt_tác của nhân_loại	pos

Figure 3. The structure of the sentiment analysis data file.

The structure of the files is as follows:

- *id* column: identification of pattern.
- *text* column: the content of opinions.
- polarity: sentiment label.
- 6. The Data\_two\_and\_more\_aspect\_sentences\_test folder contains the two\_and\_more\_aspect\_sentences\_test.xls file for test multi-aspect in the aspect detection task. Figure 4 shows the structure of this file.

1	Α	В	C	D	E	F	G	Н	1
1	id	sentences	safety	engine	price	exterior	interior	performance	overall
2	1001	cơ nội đến ngoại thất như vậy có gọi là xe độ không	no	yes	no	yes	yes	no	no
3	1002	1 CHIẾC XE HOÀN HẢO NHƯNG HÀNG GHẾ SAU LÀ 1 THẢM	no	no	yes	no	yes	no	yes
4	1003	15 tỷ mua con Sedona ngon lành hơn nhiều	no	no	yes	no	no	no	yes
5	1004	10 tỉ bỏ ra mua cái máy công nghệ của thế kỷ 19	no	yes	yes	no	no	no	no
6	1005	tội gì tầm giá này vay tiền đi mua Fortuner 27 mẫu hơi cũ nhưng như vậy lại đỡ nhàm chán phù hợp các bác trung niên thôi	no	no	yes	yes	no	no	no

**Figure 4.** The structure of the multi-aspect sentences data file.

The structure of the files is as follows:

- *id* column: identification of pattern.
- *sentences* column: the content of sentences.

- safety, engine, price, exterior, interior, performance, and overall column: determined "yes" or "no" aspect in sentence correspond.
- 7. The Opinion\_for\_summarize folder contains the opinion\_test.xls file for test multi-aspect in the aspect detection task. Figure 5 shows the structure of this file.

2	A	В		С	D	E	F	G	Н	1
1	id	opinion		safety	engine	price	exterior	interior	performance	overall
2	100		ing xem ra không thay đổi nhiều, mà nhìn vẫn quá xấu.				neg			
			cao tốc 150 km / h thì chả có ý nghĩa gì, kia Morning hay bất kể con xe nào cũng làm được. Vấn đề							
			đến khi đạt 150 km / h hết bao nhiều thời gian. 2 Bác nói vượt xe khác chỉ nhíc cái là vượt, vấn đề là							
	100	2 xe khác kia chạ	y tốc độ bao nhiêu? 3 Cùng là máy xăng 27 (Em tin chắc là cùng công nghệ chứ TOY nó chả đặt RR		pos					neu
		sản xuất động	cơ cho con Prado) tại sao em thấy con Fortuner chạy ì thế? Chả có tý nào mạnh mễ như bác nói về con							
3		Prado	27 37 76 77 77 77 77 77 77 77 77 77 77 77 77							
4	100	5 1 CHIẾC XE HO	ÀN HẢO NHƯNG HÀNG GHẾ SAU LÀ 1 THẢM HỌA CHO CÁI GIÁ 10 TY			neu	pos	neg		
5	100	6 1 kiệt tác của n	hân loại.				4.1			pos
6	100	7 1 phiếu cho Ho	nda							pos
7	100	9 1,5 tỷ mua con	Sedona ngon lành hơn nhiều.							neg
8	100	.0 10 tỉ bỏ ra mua	cái máy công nghệ của thế kỷ 19 . máy 57 v8 mà có 368 mã lực thua máy của mec 30		neg					

Figure 5. The structure of the multi-aspect sentences data file.

The structure of the files is as follows:

- *id* column: identification of pattern.
- opinion column: the content of opinions.
- safety, engine, price, exterior, interior, performance, and overall column: indicate whether to review the respective aspect or not. If there is an aspect assessment, what is the emotional pole, positive (pos), negative (neg), or neutral (neu)?

## **B.** Source code description

The Source\_code folder contains source code files in Python that implement the proposed summary model. The main function of each file specification is as follows:

- data url.py: declare global variables.
- helpers.py: the data processing functions.
- m BiLSTM multi gpu.py: the BiLSTM model
- m CNN.py: the CNN model
- m CNN BiLSTM.py: the CNN BiLSTM model
- m\_CNN\_LSTM.py: the CNN LSTM model
- m data aspect.py: the data processing for aspect functions.
- m LSTM.py: the LSTM model.
- multi opinion summarize final.py: multi-opinion summary system
- single opinion summarize final.py: single-opinion summary system.
- CNN sentiment word 3c relu.json: The configuration of the CNN model.
- CNN\_sentiment\_word\_3c-039-0.3146-0.9361.h5: The feature of the corpus that the CNN model has trained.