Matching of Japanese Listening Test Dialogues and Anime Scene Dialogues based on Zero-shot Attribute Classification

Yangdi NI, Junjie SHAN, Yoko NISHIHARA

College of Information Science and Engineering, Ritsumeikan University

Background

- Many studies use anime for teaching [Shan, 2021] [Haruka,2017]
- Using anime that has the same attribute as listening test is better for teaching [Ni, 2023]
- Matching texts with the same attribute requires classifying the text before

Task	Method	Demerit
Matching of same attribute texts	Manually labeling	Take a lot of time
	Classification model training	1.Require training date 2.Take time to train the model

Zero-shot classification method doesn't need training data and doesn't need to train the model

Purpose of research

- Using zero-shot classification to classify the attributes of the listening test and anime and match them
- Zero-shot classification has flexibility. We don’t need to train the model for every attribute as traditional classification.
- Automatic matching of listening test and anime scene with the same attribute without training classification model

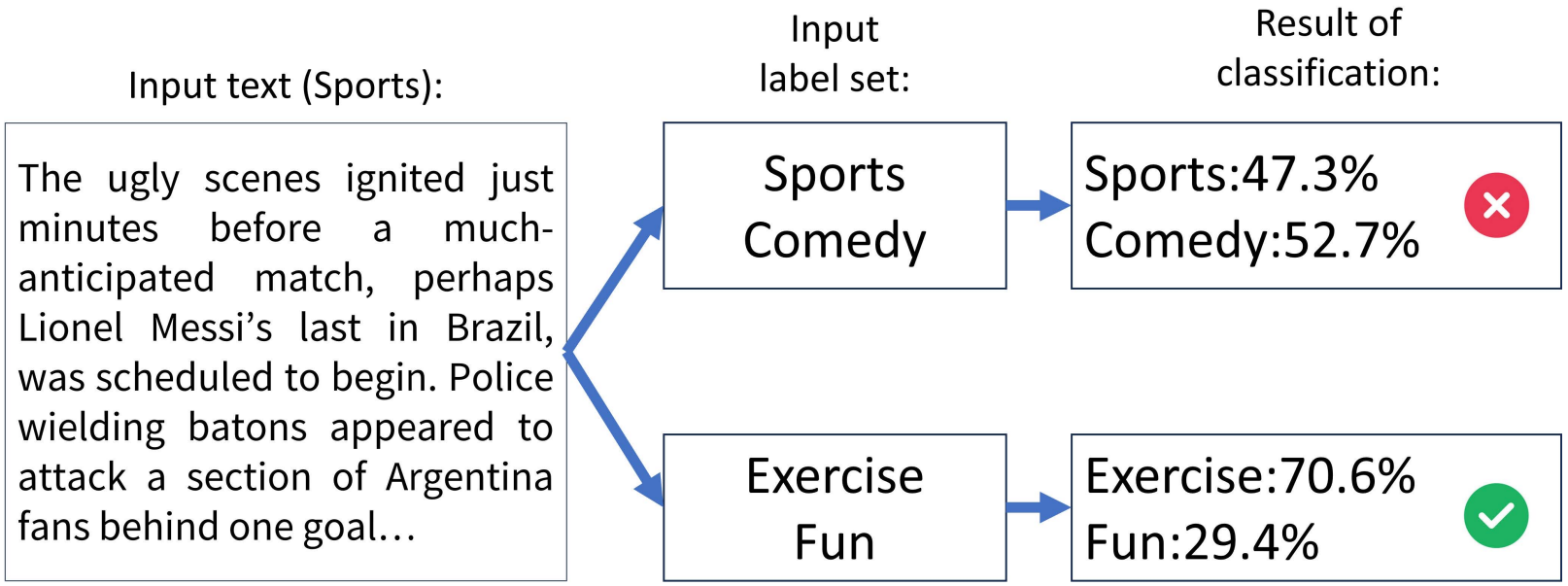
Zero-shot classification

- Zero-shot classification has two inputs:

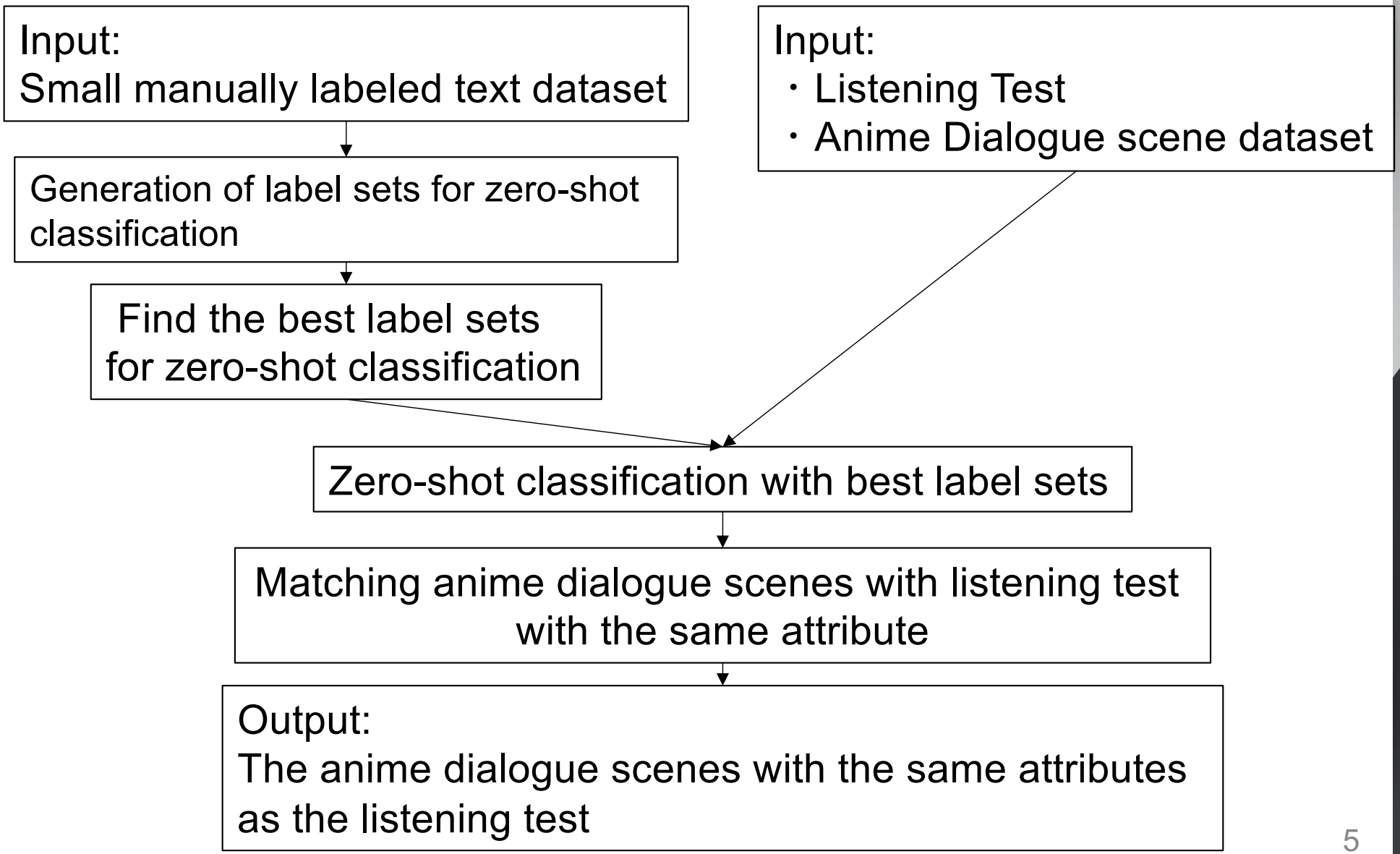
(1) Text to be classified

(2) Label sets of the classification

- For the exact input text, the effect of zero-shot classification will change when the input label to be classified changes



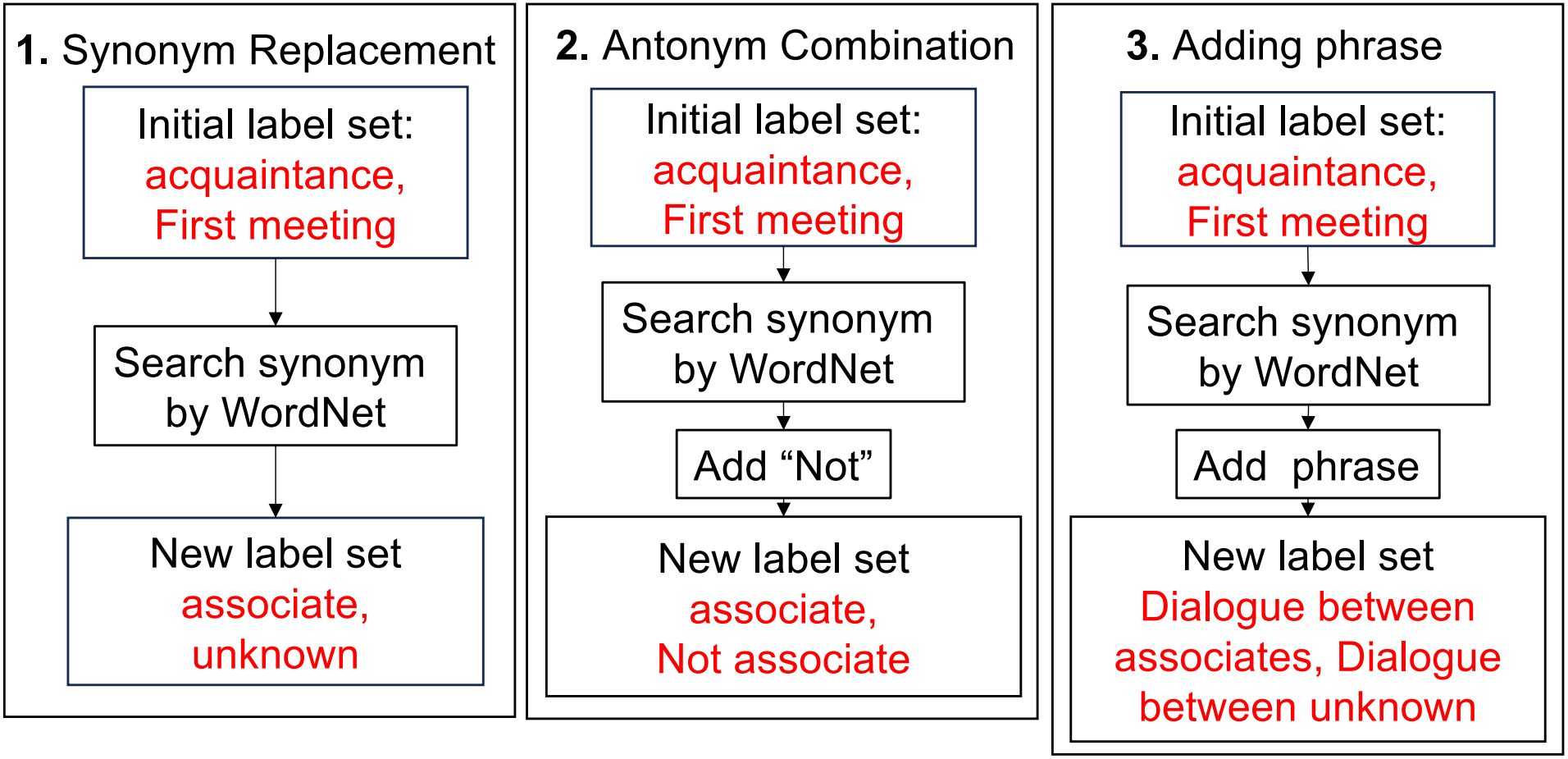
Proposed Method



Generation of label sets for zero-shot classification

Using different label sets, the effect of zero-shot classification will be different

Method to generate label set:



Adding phrase to label set can make the effect of classification better [Radford,2021]

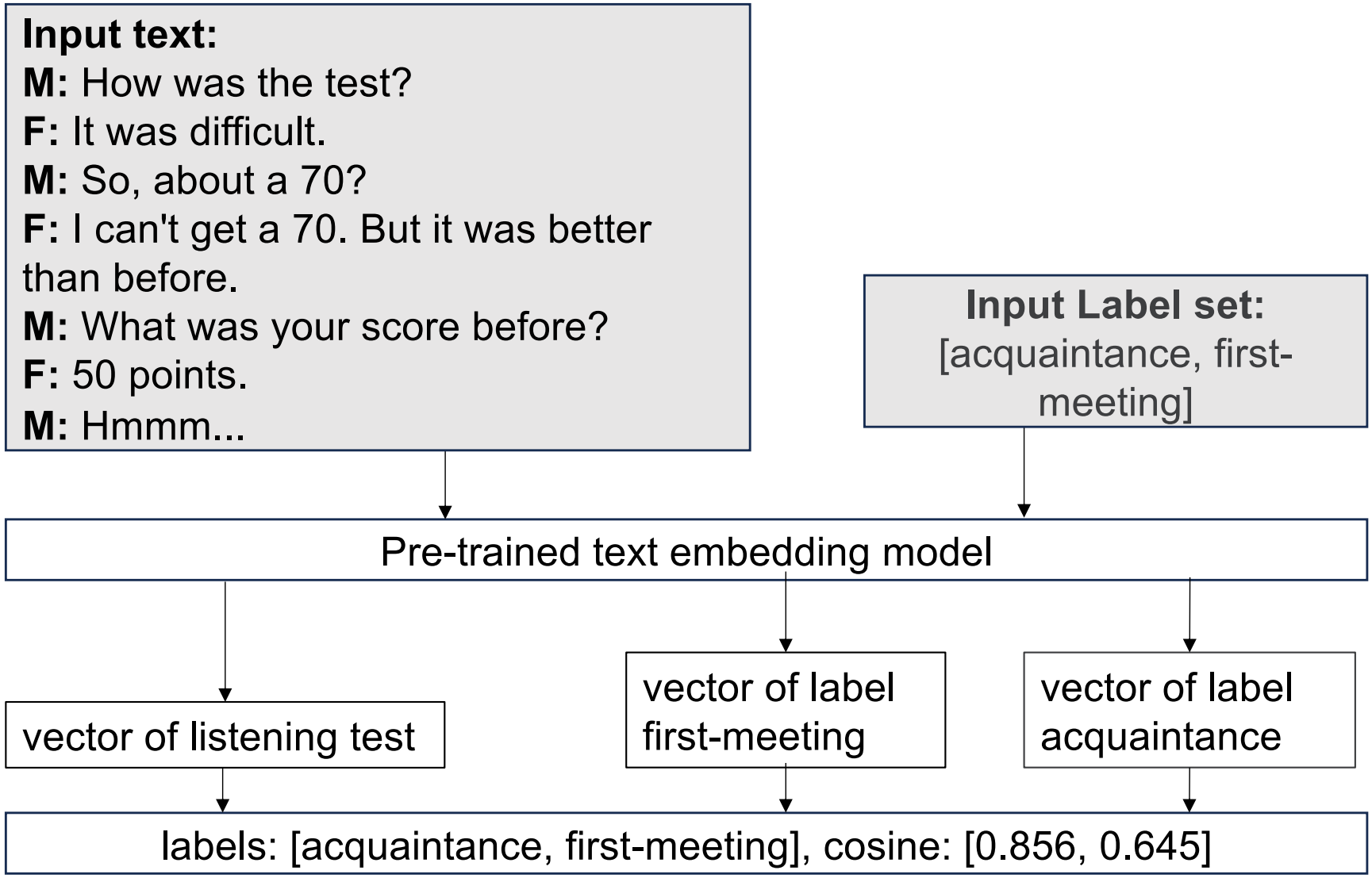
zero-shot classification method

Two zero-shot classification methods:

1.text embedding-based cosine similarity

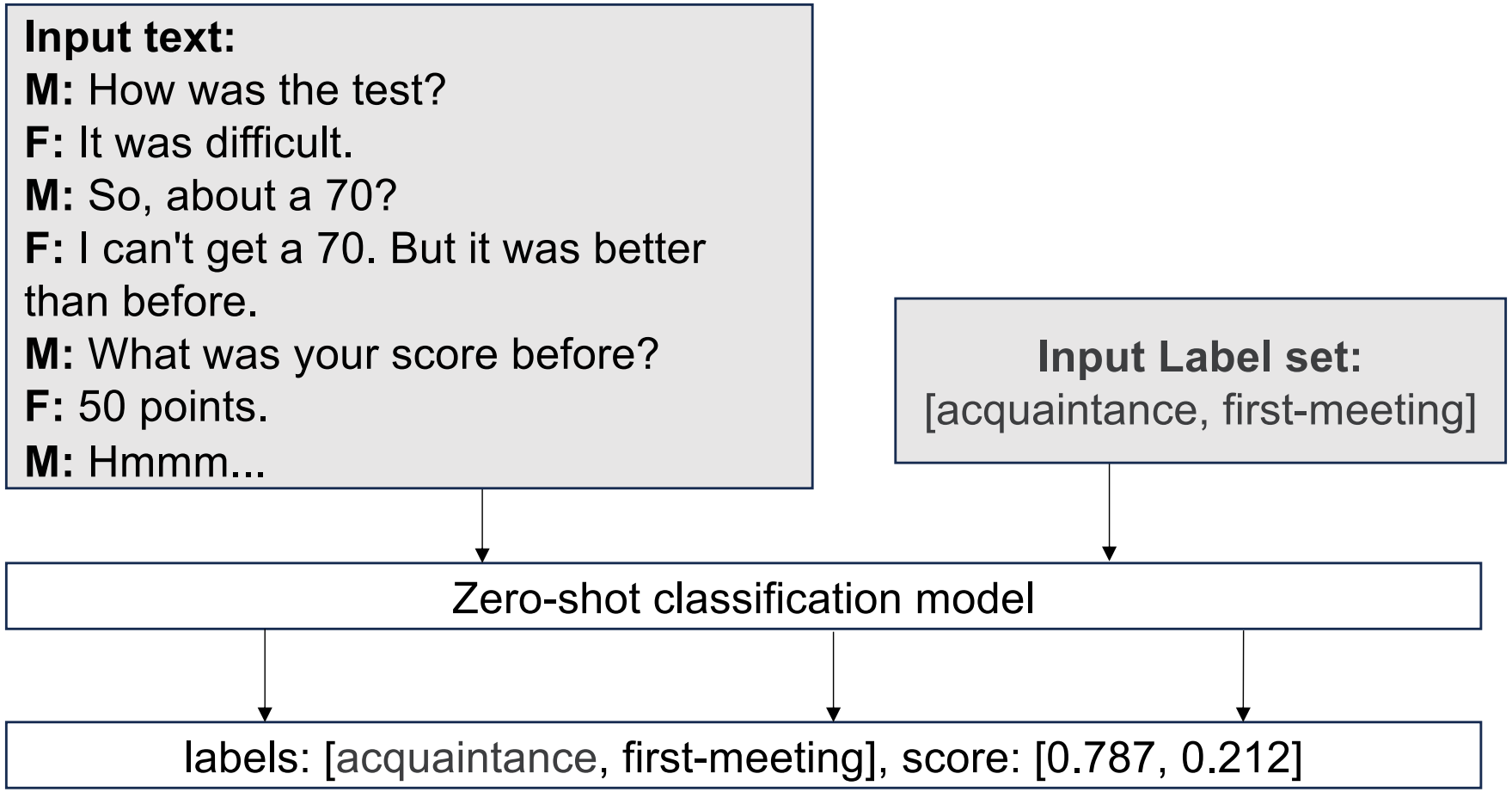
2.end-to-end pre-trained zero-shot model

Text embedding-based cosine similarity



The model used in this study: BERT-embedding

End-to-end pre-trained zero-shot model



The model used in this study: MoritzLaurer (https://huggingface.co/MoritzLaurer/mDeBERTa-v3-base-xnli-multilingual-nli-2mil7)

Zero-shot Attribute Matching

- Classify the attributes of listening test dialogues and anime dialogue scenes by zero-shot classification match
- The listening dialogues and anime dialogues with their classified attributes by three matching mode:

Inputted Listening test

Attribute:
Label1: first-meeting
Label2: discuss
Label3: on the phone

1. Single-attribute

Matched Anime scenes

Attribute:
Label1: first-meeting
Label2: --
Label3: --

2. Double-attributes

Matched anime scenes

Attribute:
Label1: first-meeting
Label2: discuss
Label3: --

3. Triple-attributes

Matched anime scenes

Attribute:
Label1: first-meeting
Label2: discuss
Label3: on the phone

Experiment (Model-Label Test)

Purpose :
Find the best label sets for zero-shot classification

Process:
1. Manually labeled 250 listening tests.

2. Generated 162 candidate label sets using three methods: Synonym Replacement, Antonym Combination, Adding phrase

3. 250 listening tests were classified by each zero-shot classification method, the classification results were compared with manual results, and the classification effect of the label set was evaluated.

4. Find the best label sets for each zero-shot classification method

Evaluation :
Root Mean Square Error (RMSE)
Hard to balance the attribute of listening tests

Listening test data used in Model-Label Test

Attribute	Label set	Number
Speaker's relationship (four categories)	first-meeting	49
	acquaintance	82
	friends	98
	family	21
Dialogue style (two categories)	chat	75
	consultation	175
Dialogue location (six categories)	home	39
	workplace	79
	school	53
	facilities	42
	outdoor	14
	telephone	23

The number of label for every attribute is imbalanced

Result of Model-Label Test

Bert-embedding

Attribute	Best label set	RMSE
Speaker's relationship (four categories)	[first meeting, acquaintance, friend, relative]	3.00
Dialogue style (two categories)	[informal, formal]	3.93
Dialogue location (six categories)	[speaking at home, speaking at work, speaking at school, speaking at public commercial facilities, speaking on the street, speaking on the phone]	3.88

Result of Model-Label Test MoritzLaurer

Attribute	Best label set	RMSE
Speaker's relationship (four categories)	[conversation between first-meeting, conversation between acquaintances, conversation between friends, conversation between relatives]	2.39
Dialogue style (two categories)	[small talk, consultation]	3.43
Dialogue location (six categories)	[speaking at home, speaking at work, speaking at school, speaking at facilities, speaking on the street, speaking on the phone]	1.75

The best label set extracted with the MoritzLaurer model is better than the best label set extracted with Bert embedding

Experiment(Zero-shot Attribute Matching)

Purpose：By using the proposed method, evaluate whether the proposed method can match the conversation of a similar anime scene to the listening test.

data：

Statistic Item	Listening Tests	Anime Scenes
Number of count	250	314,930
Min Length	37	88
Max Length	743	336
Avg. Length	211.5	189
Std. of Length	124.3	73.4

Evaluation :

Number of matched Anime Scenes

WCR: This measures the word cover rate in the sampling situation (random 100 samples, get word cover rate, average from 10 times)

Text-SIM: This measures the text similarity in the sampling situation (random 100 samples, get text similarity, average from 10 times)

Result of Zero-shot Attribute Matching

Zero-shot classification method	Matching mode	# of Matched Anime Scenes
Baseline	Random	314,930
BERT-embedding	Single-attribute	202,908
	Double-attribute	93,245
	Triple-attribute	58,908
MoritzLaurer	Single-attribute	**174,902**
	Double-attribute	**82,345**
	Triple-attribute	**47,907**

# of Matched Anime Scenes: MoritzLaurer < BERT-embedding

Result of Zero-shot Attribute Matching

Zero-shot classification method	Matching mode	WCR
Baseline	Random	69.37%
BERT-embedding	Single-attribute	69.45%*
	Double-attribute	69.70%**
	Triple-attribute	70.19%**
MoritzLaurer	Single-attribute	**69.86%****
	Double-attribute	**70.49%****
	Triple-attribute	**71.50%****

*p<0.05, **p<0.01

WCR: MoritzLaurer > BERT-embedding > Baseline

Result of Zero-shot Attribute Matching

Zero-shot classification method	Matching mode	Text-SIM
Baseline	Random	0.8184
BERT-embedding	Single-attribute	**0.8247***
	Double-attribute	**0.8314****
	Triple-attribute	**0.8432****
MoritzLaurer	Single-attribute	0.8219**
	Double-attribute	0.8229**
	Triple-attribute	0.8245**

*p<0.05, **p<0.01

Text-SIM: BERT-embedding > MoritzLaurer > Baseline

Conclusion

- In this study, we propose a method for matching attributes of listening tests and anime dialogue scenes using zero-shot classification.
- The proposed method generates the best label sets for zero-shot classification, classifies listening tests and anime dialogue scenes according to their attributes, and matches them.
- Evaluation experiments confirmed that the proposed method can match anime dialogue scenes with listening tests that have similar contexts.

Thanks For Listening