THE 12TH INTERNATIONAL SYMPOSIUM ON INFORMATION AND COMMUNICATION TECHNOLOGY

DECEMBER 7-8, 2023 HO CHI MINH CITY, VIETNAM



Dialogue Attributes' Zero-Shot Classification **Based Anime Scene's Matching for Japanese Listening Test**

Junjie SHAN

Yangdi NI

Yoko NISHIHARA

shan@fc.ritsumei.ac.jp is0598eh@ed.ritsumei.ac.jp Ritsumeikan Global Innovation Research Organization Kusatsu, Shiga, Japan

nisihara@fc.ritsumei.ac.jp

Introduction

Traditional classification models:

- 1. Need massive manually annotated training data.
- 2. Need many training source for model training
- 3. Are not generalized well and task-specified.

Zero-shot Classification:

- 1. No-need training data, good generalization
- 2. Even for a same input text, when inputted label set is changed, the zero-shot classification result also may change as well

Input text:

The ugly scenes ignited just minutes before muchanticipated match, perhaps Lionel Messi's last in Brazil, was scheduled to begin. Police wielding batons appeared to attack a section of Argentina fans behind one goal...

initial label

sets:

synonymous terms:

new label sets:

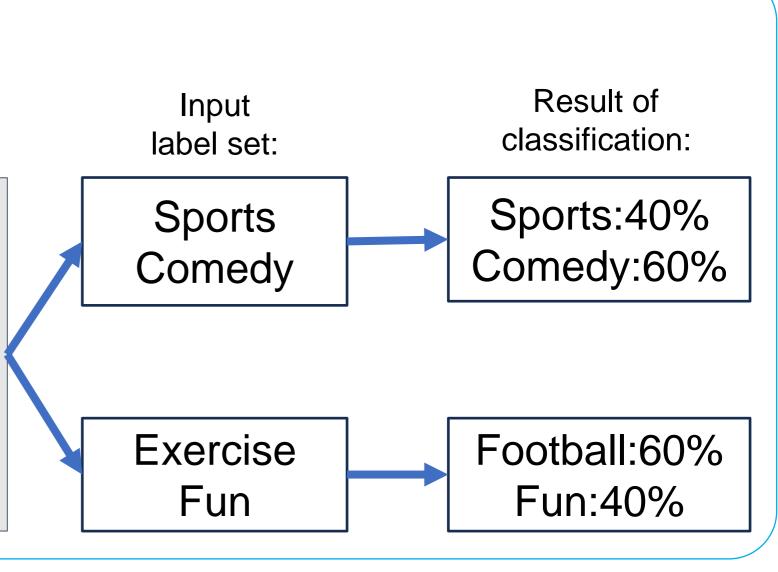
hot Attribute

Matching

chitchat, chin-wag, ...

chin-wag; discussion

chitchat; discussion



gossip; consultation

WordNet

synonyms

The flow of method

- 1. Manually label 90 JLPT listening tests
- 2. Use WordNet to Prepare 2143 label keyword sets
- 3. Use label sets and zeroshot classification model to categorize listening tests
- 4. Select the top 9 label sets with the highest accuracy as "best label set"

Input:

Listening Test

Dialogues

Listening Test Dialogues Zero-shot

Manually Annotate Attributes

Zero-shot Label Test

Test Anime Dialogue Best Label Set scenes

Label Keyword

Attribute Zero-shot Classification

Attribute Label Matching

Matched Anime Dialogue Scenes Output:

5.Use the "best label set" to zero-shot categorize the dialogue attributes of the

discussion, advice, ...

combination

chin-wag, advice

chitchat; advice

listening tests. 6. Match anime scenes and listening tests based categorized attributes

collected anime scenes and

7. Compare the matched anime scenes with the input listening test dialogues

Results

- 1.# of Anime Scenes: Matched anime scenes amount.
- 2.WCR (Sampling): Word Cover Rate with a sampling method.
- 3.Text-SIM (Sampling): Text similarity with a sampling method.
- 4.WCR (No-Sampling): Word Cover Rate with all matched anime scenes.
- 5.WCR-Eff (No-Sampling): Word Cover Rate Efficiency with all matched anime scenes.

Three matching three modes:

- 1.Single-Attribute: use only one original attribute type to match
- 1.Double-Attribute: Combines the original three attribute types two by two, use two attribute types to match
- 1.Triple-Attribute: use all three attribute types to match

Mode	# of Anime Scenes	WCR (Sampling)	Text-SIM (Sampling)	WCR(No- Sampling)	WCR-Eff(No- Sampling)
Baseline (No Match)	247,645	65.81%	0.8204	97.10%	3.92E-06
Single- Attribute	140,988	66.43%*	0.8215*	96.18%	6.82E-06
Double- Attribute	70,834	67.2%**	0.8225**	95.22%	1.34E-05
Triple-	20 262	67 7006**	U 8338**	04 5106	2 225 05

(*p<0.05,**p<0.01)

67.79%** 0.8238**

ORGANIZERS

29,362







Attribute







3.22E-05

94.51%

SPONSORS







SAMSUNG