The Cross-Lingual Conversation Summarization Challenge

Yulong Chen, Ming Zhong, Xuefeng Bai, Naihao Deng, Jing Li, Xianchao Zhu, Yue Zhang

Zhejiang University
School of Engineering, Westlake University
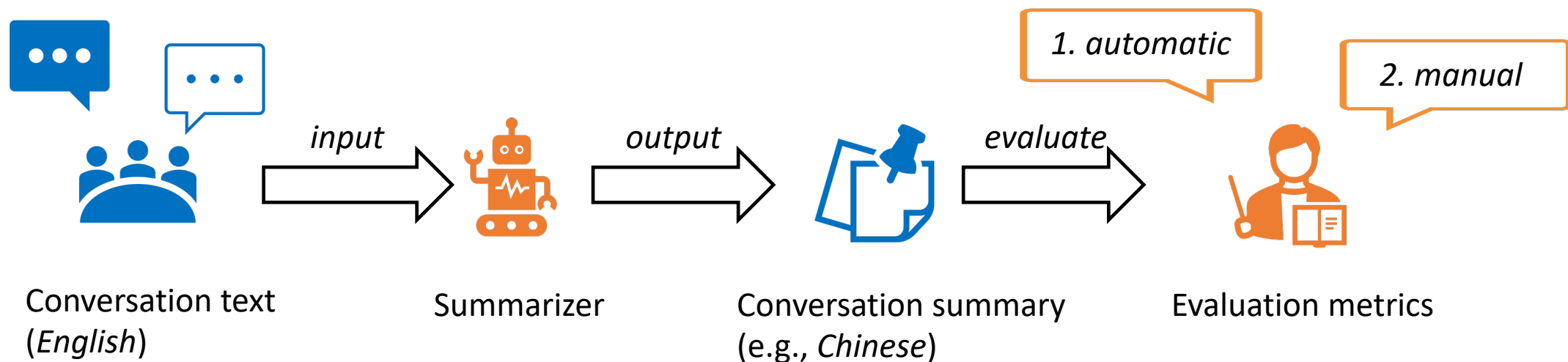**University of Illinois at Urbana-Champaign**
University of Michigan, Ann Arbor
Sichuan Lan-bridge Information Technology Co., Ltd

The *ConvSumX Challenge*

Task Overview

The cross-lingual conversation summarization (**ConvSumX**) task asks a model to *generate a summary* in *a target language* given *a piece of dialogue text* in *a source language*.

Introduction

Why Cross-lingual Conversation Summarization?
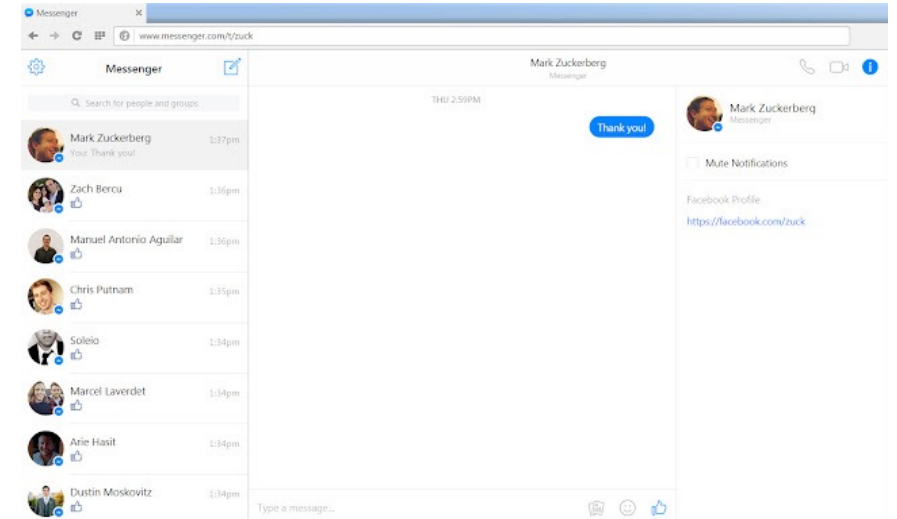
Existing conversation summarization researches focus on *English* texts.

(a) real-life scenario (DialogSum)   (b) meeting (QMSum)   (c) online chatting (SAMSum)

Introduction

Why Cross-lingual Conversation Summarization?

Existing conversation summarization researches focus on *English* texts.

A potential barrier for Non-English speakers

(a) real-life scenario (DialogSum)   (b) meeting (QMSum)   (c) online chatting (SAMSum)

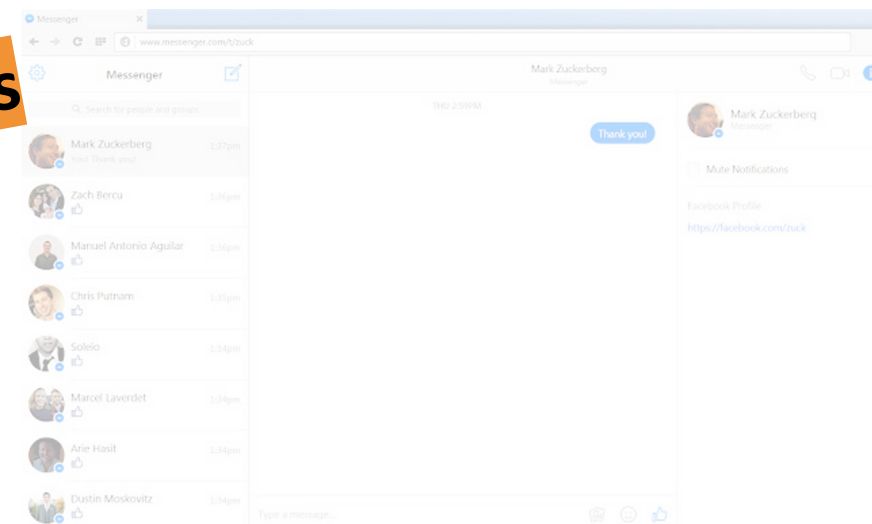# Introduction

Why Cross-lingual Conversation Summarization?

Existing conversation summarization researches focus on *English* texts.

A potential barrier for Non-English speakers

How can non-English speakers understand them?

(a) real-life scenario (DialogSum)   (b) meeting (QMSum)   (c) online chatting (SAMSum)

Introduction

Potential Solutions to Cross-lingual Conversation Summarization

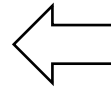This problem can be potentially solved by 3 existing methods

- First translate, then summarization pipeline

- First summarization, then translate pipeline

- End-to-end method

Introduction

Potential Solutions to Cross-lingual Conversation Summarization

This problem can be potentially solved by 3 existing methods

- First translate, then summarization pipeline

- First summarization, then translate pipeline ⟵ **Error propagation caused by machine translation and summarization systems**

- End-to-end method

Introduction

Potential Solutions to Cross-lingual Conversation Summarization

This problem can be potentially solved by 3 existing methods

- First translate, then summarization pipeline

- First summarization, then translate pipeline ⟵ **Error propagation caused by machine translation and summarization systems**
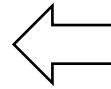
- End-to-end method

Introduction

The ConvSumX challenge

- Both personal and business need

- a general domain/language transfer technology

- NLP for social good

Task Description

ConvSumX: 2 tracks, 3 language directions

Daily-life Dialogue

#Person_1#: Good morning. I wonder whether you have got an answer from your superior.
#Person_2#: Yes, we had a meeting about it yesterday afternoon.
#Person_1#: What's the answer?
......
#Person_1#: I thought you would. So I rang Auckland last night. As you are our biggest customer, they agreed to ship the order on the first vessel available that will leave Auckland next month.
#Person_2#: Good, if you agree we'll draft the agreement right away and sign it then.
#Person_1#: By all means.

En
#Person_1# and #Person_2# agree to sign an agreement since #Person_1# could speed up the delivery as #Person_2# hopes.

Zh
由于#Person_1#将会按照#Person_2# 所期望的提升交付速度，#Person_1#和#Person_2#达成一致签署合同。

Fr
#Person_1# et #Person_2# acceptent de signer un contrat étant donné que #Person_1# est en mesure d'accélérer le processus de livraison comme souhaité par #Person_2#.

Uk
#Person_1# i #Person_2# погоджуються підписати угоду, оскільки #Person_1# може прискорити доставку, як бажає #Person_2#.

(a)

Product Meeting Transcript

Turn 0: Project Manager: We have been provided with some technical tools to communicate.
......
Turn 16: **Marketing: This is just a presentation on the trends that we're gonna use to make the product stand out from ......**
......
Turn 85: **Marketing: What do you think of adding an LCD?**
......
Turn 89: **Project Manager: Okay, we'll include it to make the appearance attractive to young people.**
......
Turn 316: Project Manager: Thanks. Have a nice day!

En
query — Summarize the discussion about the trends of current remote controls.
summary — The group discussed different trends based on different ages of people. ...... Finally they decided to add LCD screen.

Zh
query — 总结关于目前遥控器发展趋势的讨论。
summary — 根据不同年龄的人，该组讨论了不同趋势。... 最后，他们决定添加液晶显示器。

Fr
query — Veuillez écrire un compte-rendu de la discussion sur les tendances des télécommandes actuelles.
summary — Les discussions du groupe ont porté sur les différentes tendances en fonction de l'âge des personnes. ...... Ils ont finalement décidé d'ajouter un écran LCD.

Uk
query — Підведіть підсумок дискусії про тенденції сучасного дистанційного керування.
summary — Група обговорила різні тенденції, засновані на різному віці людей. ...... Нарешті вони вирішили додати рідкокристалічний екран.

(b)

Data Construction

Data selection

Track	Data Source	Domain	Query	# Conv.	# Summ.	Train/Dev/Test
Track 1	DIALOGSUM	Daily-life Dialogue	✗	131.0	13.8	12,460/500/500
Track 2	QMSum	Product Meeting	✓	6,007.7	70.5	690/145/151
		Academic Meeting	✓	13,317.3	52.7	259/54/56

Data Construction

Annotation

Gold dev & test sets
Each summary is annotated into **3** target languages by **expert annotators**.

Data Construction

Annotation

Gold dev & test sets

       Each summary is annotated into **3** target languages by **expert annotators**.

Data Construction

Annotation

Gold dev & test sets

Each summary is annotated into **3** target languages by **expert annotators.**

Silver training set

- Google translate
- NiuTrans
- LanMT

Data Construction

Annotation

Gold dev & test sets

Each summary is annotated into **3** target languages by **expert annotators.**

Silver training set

- Google translate
- NiuTrans
- LanMT

Evaluation

Evaluation contains both automatic and human evaluation

Automatic Evaluation
Rouge Scores

Human Evaluation
Combining evaluation metrics from both machine translation and summarization
- Fluency and use of language
- Relevance
- Factual and translation consistency
- Terminology
- Overall score

Evaluation

Evaluation contains both automatic and human evaluation

Automatic Evaluation
  Rouge Scores

Human Evaluation
  Combining evaluation metrics from both machine translation and summarization
  - Fluency and use of language
  - Relevance
  - Factual and translation consistency
  - Terminology
  - Overall score

Protocol

We propose the following schedule for ConvSumX

Timeline	Schedule
Jul 2022 – Dec 2022	Phase1: The task is announced, call for participant.
Dec 2022 – Mar 2023	Phase2: Leaderboard is open.
Mar 2023 – Jun 2023	Phase3: Submission is closed; begin human evaluation

Protocol

We propose the following schedule for ConvSumX

Timeline	Schedule
Jul 2022 – Dec 2022	Phase1: The task is announced, call for participant.
Dec 2022 – Mar 2023	Phase2: Leaderboard is open.
Mar 2023 – Jun 2023	Phase3: Submission is closed; begin human evaluation

Thank you!