InstructABSA: Instruction Learning for Aspect Based Sentiment Analysis

Kevin Scaria, Himanshu Gupta, Saurabh Arjun Sawant, Swaroop Mishra, Chitta Baral

Arizona State University

Introduction

Aspect-Based Sentiment Analysis은 사용자 표현에서 세부적인 감정을 이해하는 데 중요한 작업

ABSA는 아래 그림과 같이 화자의 의견을 이해함으로써 감정의 극성을 분류하는 작업

Subtask	Input	Output
Aspect Term Extraction (ATE)	S_i	a^1, a^2
Aspect Term Sentiment Classification (ATSC)	$S_i + a^1, S_i + a^2$	sp^1, sp^2
Joint Task	S_i	$(a^1, sp^1), (a^2, sp^2)$

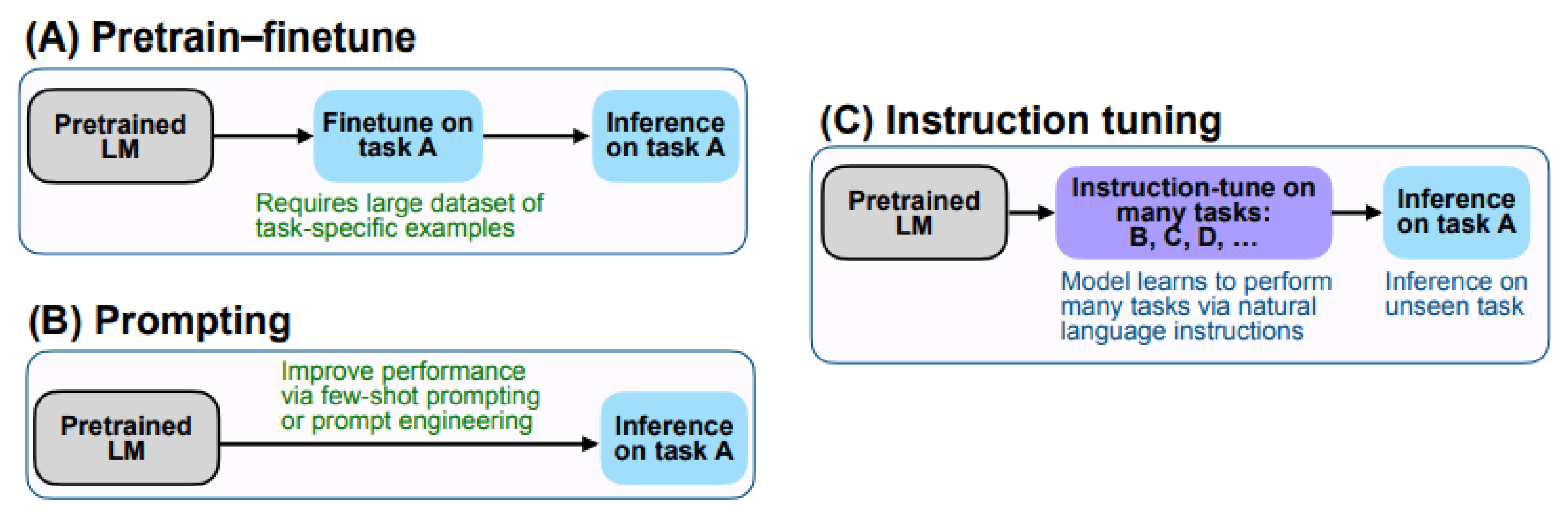
Figure 1: Illustration of the three ABSA subtasks where S_i is the i^{th} sentence, a^i is the aspect terms, sp^i is the sentiment polarity and o^i is the opinion term.

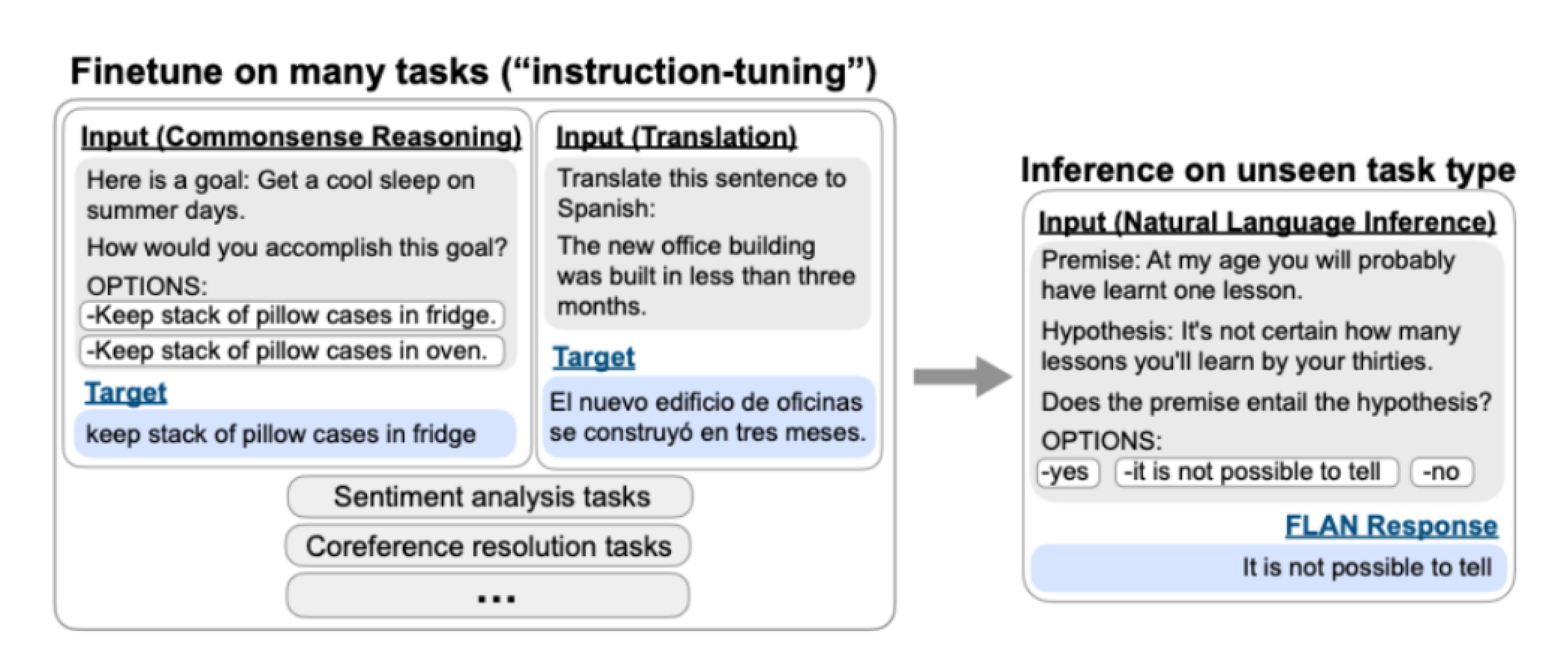
Introduction

- Encoder, Decoder
- Transformer-based models
  → limitations like information loss and ignoring semantic labels
- Instruction learning

Introduction

- Instruction learning



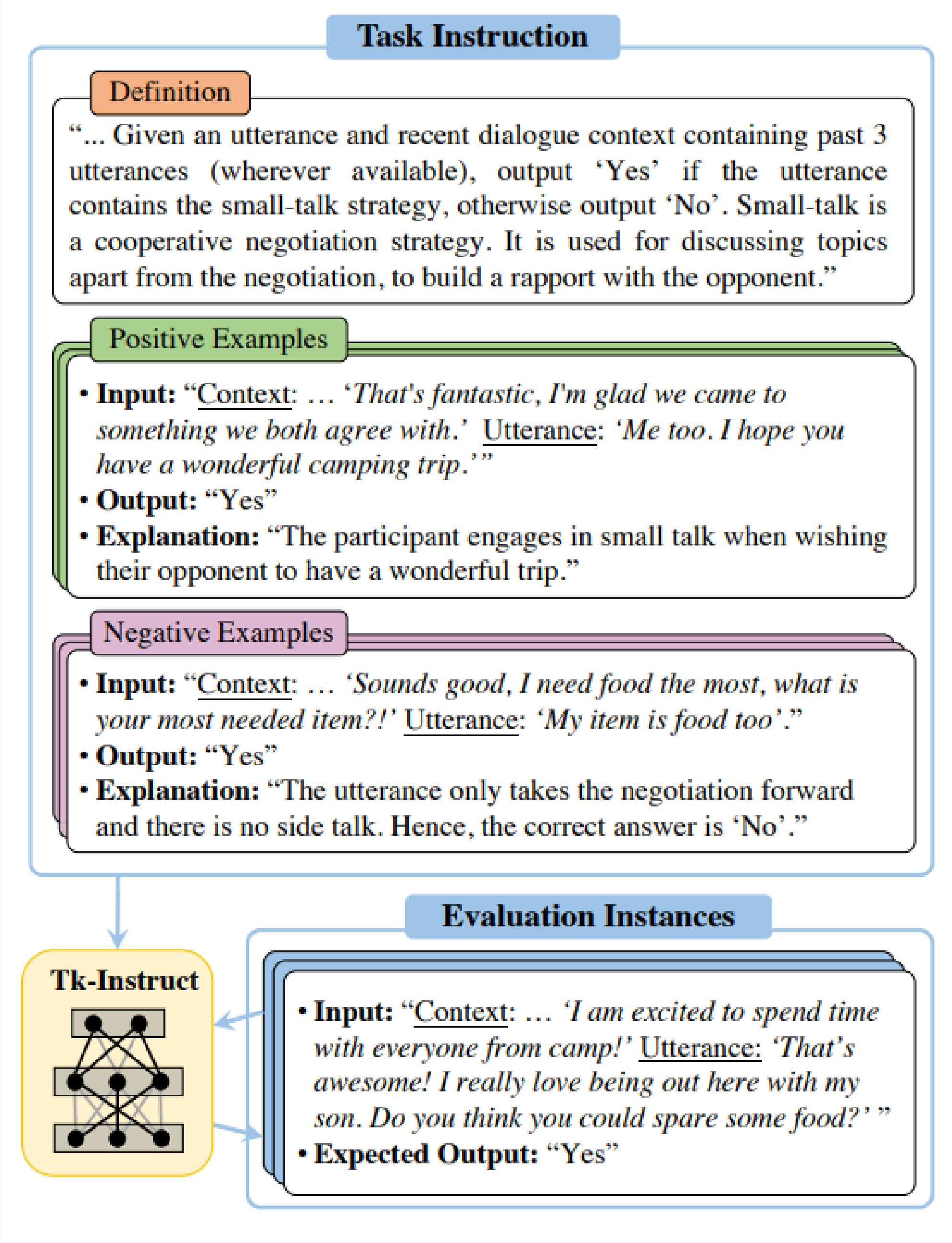


여러 task를 포함한 데이터셋을 통해 finetuning을 수행

모델에게 일종의 명령/지침으로 표현된 NLP Task를 수행하는 방법을 가르쳐서 자연어를 처리하고 이해하는 능력을 향상

Introduction

- Tk-Instruct model(Task Instruction)



DEFINITION defines a given task in natural language

POSITIVE EXAMPLES are samples of inputs and their correct outputs

NEGATIVE EXAMPLES are samples of inputs and their incorrect/invalid outputs

InstructABSA: Instruction Learning

Aspect Term Extraction (ATE)

- $A_i = LM_{ATE}(S_i)$ $S_i = \{w_i^1, w_i^2, ... w_i^n\},$:문장내의 토큰의 갯수 $A_i = a_i^1, a_i^2 .., a_i^m$: 문장내의 속성어의 갯수

Aspect Term Sentiment Classification (ATSC)

- $sp_i^k = LM_{ATSC}(S_i, a_i^k)$ sp_i^k 는 감성어 [긍정,부정,중립]을 의미

Joint Task

- $[A_i, SP_i] = LM_{Joint}(S_i)$

Proposed Approach

Aspect Term Extraction (ATE)

- $A_i = LM_{Inst}(Inst, S_i)$

Aspect Term Sentiment Classification (ATSC)

- $sp_i^k = LM_{Inst}(Inst, S_i, a_i^k)$

Joint Task

- $[A_i, SP_i] = LM_{Inst}(Inst, S_i)$

Task	Aspect Term Extraction (ATE)
Definition	Definition: The output will be the aspects (both implicit and explicit) which have an associated opinion that is extracted from the input text. In cases where there are no aspects, the output should be noaspectterm.
Positive Example	Example Input 1: With the great variety on the menu, I eat here often and never get bored. Example Output 1: menu Example Input 2: Great food, good size menu, great service, and an unpretentious setting. Example output 2: food, menu, service, setting
Negative Example	Negative input 1: They did not have mayonnaise, forgot our toast, left out ingredients (ie, cheese in an omelet), below hot temperatures and the bacon was so overcooked it crumbled on the plate when you touched it. Negative output 1: toast, mayonnaise, bacon, ingredients, plate Negative input 2: The seats are uncomfortable if you are sitting against the wall on wooden benches. Negative output 2: seats
Neutral Example	Neutral Input 1: I asked for a seltzer with lime, no ice. Neutral Output 1: seltzer with lime Neutral Input 2: They wouldn't even let me finish my glass of wine before offering another. Neutral Output 2: glass of wine
Input	Now complete the following example- input: My son and his girlfriend both wanted cheeseburgers and they were huge! output: cheeseburgers

Table 11: Illustrating InstructABSA-2 instruction prompting for the ATE sub task.

Experimental Setup

Base model: Tk-Instruct-base-def-pos as the instruction tuned model LM_{Inst}

InstructABSA-1 : Definition , 2 positive examples

InstructABSA-2 : Definition , 2 positive, negative, and neutral examples

Dataset: SemEval 2014 Task-4 Lapt14,Rest14 cross-domain experiments

Evaluation Metric: F1-score (ATE,joint task) , accuracy(ATSC)

Results - ATE

Model	Lapt14	Rest14
$GPT2_{med}$ (SB1&2)	81.52	75.94
$GPT2_{med}$ (SB1&4)	82.04	-
GRACE	87.93	85.45
InstructABSA-1	$91.40_{\uparrow 3.47}$	**92.76**$_{\uparrow 7.31}$
InstructABSA-2	**92.30**$_{\uparrow 4.37}$	$92.10_{\uparrow 6.65}$

Table 1: ATE subtask results denoting F1 scores. $GPT2_{med}$ (SB1&2 and SB1&4) and GRACE results are from Hosseini-Asl et al. (2022) and Luo et al. (2020).

InstructABSA uses a model with 200M parameters

GPT2 uses a model with 1.5B parameters

Results - ATSC

Model	Lapt14	Rest14
ABSA-DeBERTa	82.76	89.46
LSAS-XdeBERTa	86.21	90.33
LSAT-XdeBERTa	86.31	**90.86**
InstructABSA-1	**88.37**$_{\uparrow 2.06}$	87.42$_{\downarrow 3.44}$
InstructABSA-2	85.85$_{\downarrow 0.46}$	89.76$_{\downarrow 1.10}$

Table 2: ATSC subtask results denoting accuracy. ABSA-DeBERTa and XdeBERTa (LSAT and LSAS) results are from Marcacini and Silva (2021) and Yang and Li (2021), respectively.

InstructABSA uses a model with 200M parameters

XdeBERTa uses a model with 355M parameters

Results - Joint Task

Model	Lapt14	Rest14
GPT2_{med} (SB1&2)	53.55	60.07
SPAN	68.06	74.92
GAS-Extraction	68.64	76.58
GAS-Annotation	68.06	77.13
GRACE	70.71	78.07
InstructABSA-1	$78.89_{\uparrow 8.18}$	$76.16_{\downarrow 1.91}$
InstructABSA-2	$\mathbf{79.34}_{\uparrow 8.63}$	$\mathbf{79.47}_{\uparrow 1.40}$

Table 3: Results of the Joint Task denoting F1 scores. SPAN and GAS (Extraction and Annotation) results are from Hu et al. (2019) and Zhang et al. (2021)

InstructABSA uses a model with 200M parameters

GPT2 uses a model with 1.5B parameters

LSA uses a model with 355M parameters

Span uses a model with 345M parameters

Conclusion

InstructABSA를 사용하여 기존의 모델들 보다 작은 사이즈임에도 불구하고 높은 성능을 기록하였고

cross-domain , joint-domain의 방법을 사용하여 성능을 평가하였습니다

그러나 한계점으로는 Baseline 모델인 Tk-instruct model이 영어로 학습되어 다른 언어에서는 성능이 나오지 않는다는점이 있습니다