Two-Level Classification Using Recasted Data For Low Resource Settings

Shagun Uppal[1], Vivek Gupta[2], Avinash Swaminathan[3], Haimin Zhang[4], Debanjan Mahata[4], Rakesh Gosangi[4], Rajiv Ratn Shah[1,4] and Amanda Stent[4]

[1]IIIT-Delhi, [2]University of Utah
[3]NSUT, [4]Bloomberg

INTRODUCTION

- To address the scarcity of data in low-resource languages, we use existing classification datasets to create NLI datasets by recasting.
- We propose weakly-supervised constraints (with paired level supervision) to remove inconsistencies in the Textual Entailment (TE) predictions.

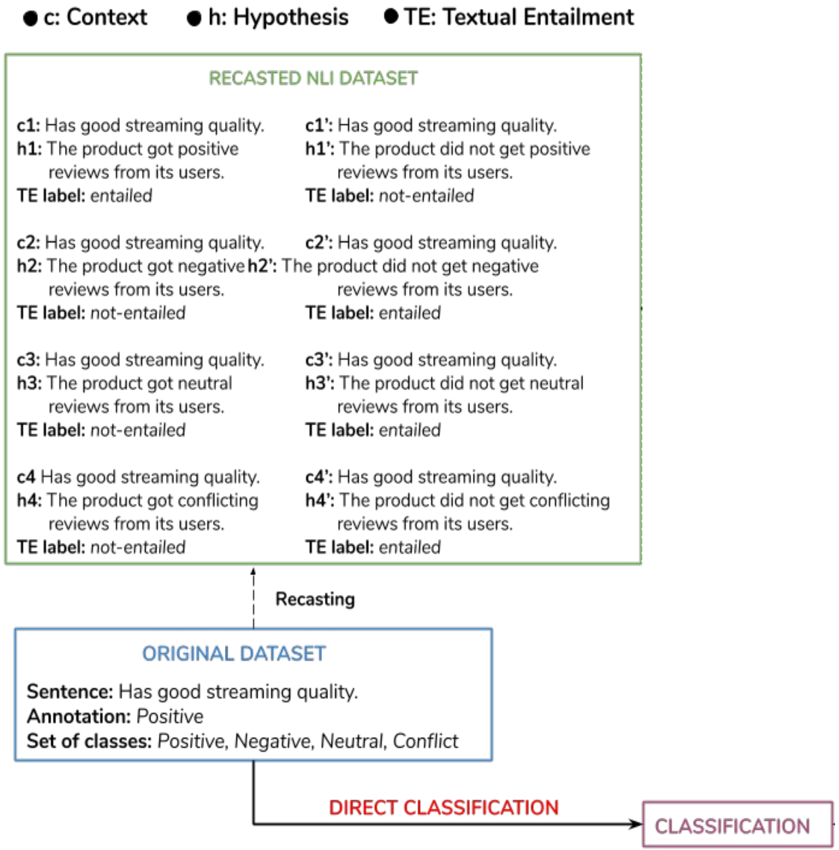
- We further use TE predictions for the classification task, with the aim to compensate for the lack of enough labelled classification data.
- This two-step classification makes it more interpretable to analyse the model understanding of reformulated language inputs.

Recasting Data

ORIGINAL DATASET

Sentence: Has good streaming quality.
Annotation: *Positive*
Set of classes: *Positive, Negative, Neutral, Conflict*

Direct Classification



RECASTING DATA

- c: Context
- h: Hypothesis
- TE: Textual Entailment

RECASTED NLI DATASET

c1: Has good streaming quality.
h1: The product got positive reviews from its users.
TE label: *entailed*

c1': Has good streaming quality.
h1': The product did not get positive reviews from its users.
TE label: *not-entailed*

c2: Has good streaming quality.
h2: The product got negative reviews from its users.
TE label: *not-entailed*

c2': Has good streaming quality.
h2': The product did not get negative reviews from its users.
TE label: *entailed*

c3: Has good streaming quality.
h3: The product got neutral reviews from its users.
TE label: *not-entailed*

c3': Has good streaming quality.
h3': The product did not get neutral reviews from its users.
TE label: *entailed*

c4 Has good streaming quality.
h4: The product got conflicting reviews from its users.
TE label: *not-entailed*

c4': Has good streaming quality.
h4': The product did not get conflicting reviews from its users.
TE label: *entailed*

Recasting

ORIGINAL DATASET

Sentence: Has good streaming quality.
Annotation: *Positive*
Set of classes: *Positive, Negative, Neutral, Conflict*

APPROACH: TEXTUAL ENTAILMENT

Textual Entailment (TE).

To analyse if the model can **draw reasonable inferences** from the **context to hypothesise** over other related/unrelated data

Textual Entailment Example

Context-Hypothesis	Label
p : The kid exclaimed with joy. h : The kid is happy.	*entailed*
p : I am feeling happy. h : I am angry.	*not-entailed (contradictory)*
p : Suzan lives in Japan. h : Suzan was born in Australia.	*not-entailed (neutral)*

Consistency Regularisation (CR).

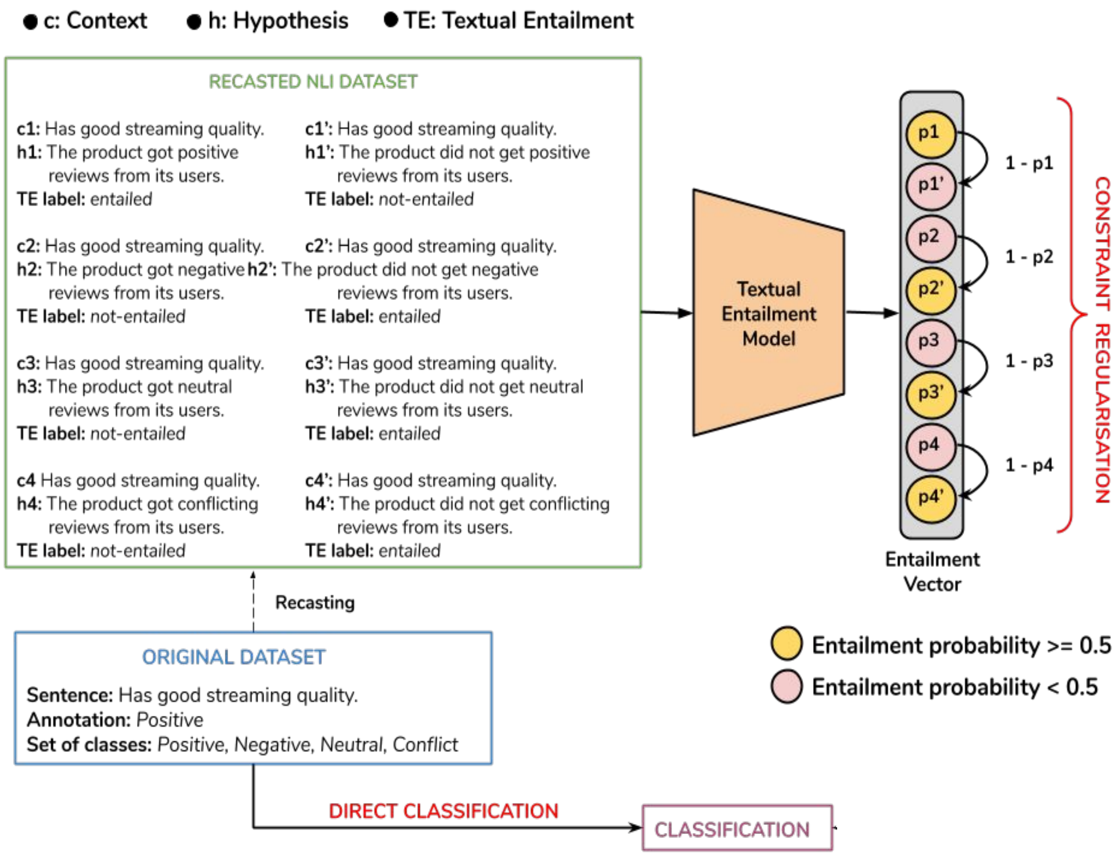
For any given **context-hypothesis pair *P***, there exists **another such pair *P'***, with **negated hypothesis and flipped TE label**.

We leverage this to **ensure a pairwise consistency** in the entailment predictions **for *P* and *P'***, such that they are always complements of each other.

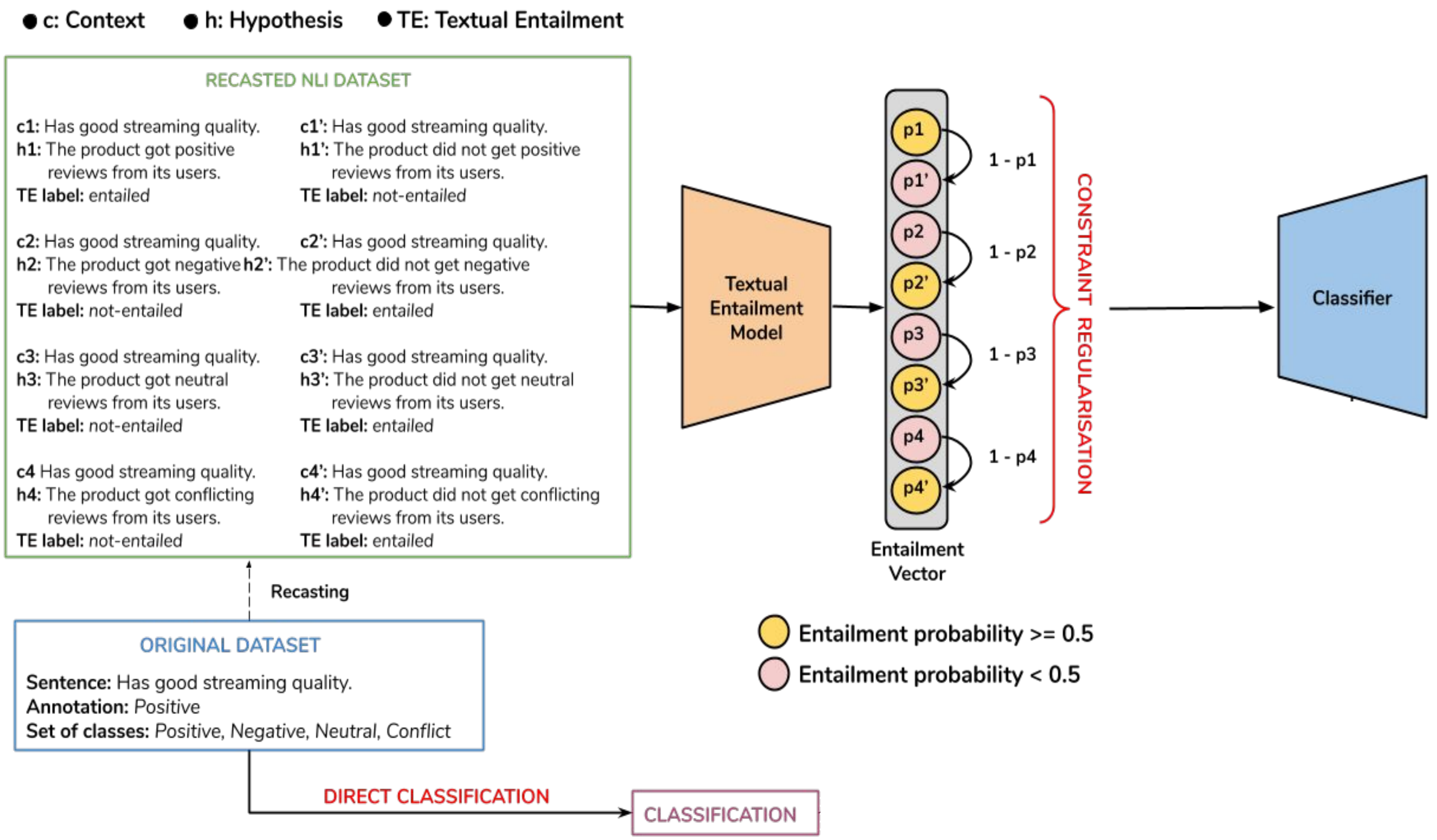
$$\mathcal{L}_{reg} = ||\mathcal{T}(P) + \mathcal{T}(P') - 1_2||^2$$

Here, $\mathscr{T}$ represents the textual entailment network.

TEXTUAL ENTAILMENT



TEXTUAL ENTAILMENT



InConsistency Example

Context (Hindi): वह रोया जब उसने अपना पालतू खो दिया **(English):** He cried over his lost pet.			**Emotion class (Hindi):** दुख **(English):** Sad	
Hypothesis (Hindi)	**Hypothesis (English)**	**TE label**	**Consistency**	**Prediction**
$h1$: वह खुश है $h1'$: वह खुश नहीं है	$h1$: He is happy. $h1'$: He is not happy.	*not-entailed* *entailed*	Consistent	Correct Correct
$h1$: वह खुश है $h1'$: वह खुश नहीं है	$h1$: He is happy. $h1'$: He is not happy.	*not-entailed* *not-entailed*	Inconsistent	Correct Incorrect
$h1$: वह खुश है $h1'$: वह खुश नहीं है	$h1$: He is happy. $h1'$: He is not happy.	*entailed* *entailed*	Inconsistent	Incorrect Correct
$h1$: वह खुश है $h1'$: वह खुश नहीं है	$h1$: He is happy. $h1'$: He is not happy.	*entailed* *not-entailed*	Consistent	Incorrect Incorrect

APPROACH: TWO-WAY CLASSIFICATION

Two-step Classification.

We extend the binary classification knowledge from **TE to a multi-class classification paradigm** to achieve two-step classification.

1. Obtain TE predictions for all re-casted augmentations of any given sentence.
2. Use these predictions to find the boundary for multi-class classification decision, as shown below.

Context: He cried over his lost pet.	
Recasted hypothesis	**Binary output**
1. He is happy.	not-entail
2. He is not happy.	entail
3. He is angry.	not-entail
4. He is not angry.	entail
5. He is sad.	entail
6. He is not sad.	not-entail
Emotion Annotation: *Sad*	

Joint Objective (JO) .

The joint end-to-end training objective (instead of independent training of TE and two-step classification) is to create a feedback between 1. and 2.

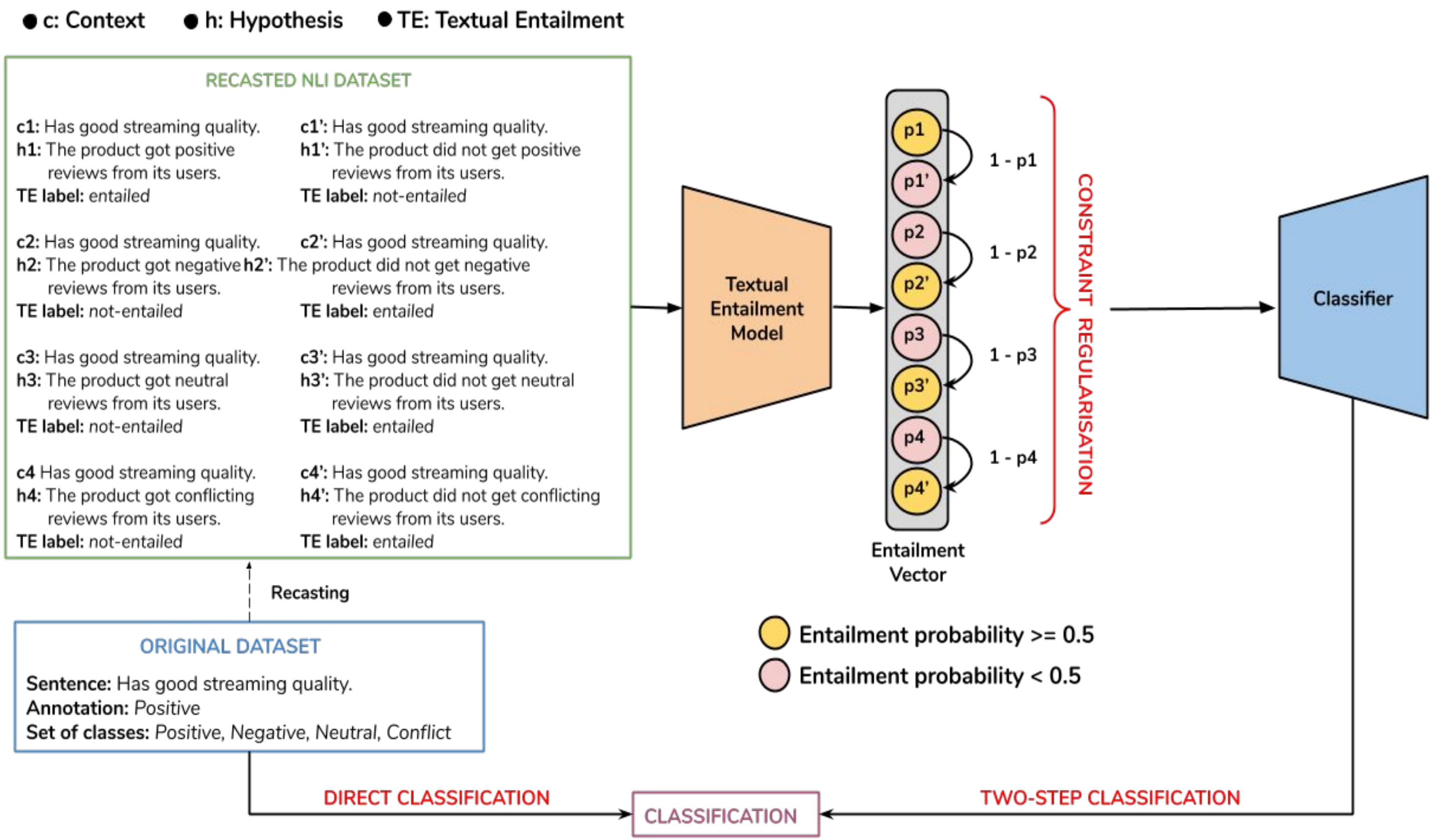
This prevent the Textual Entailment Model from acting as bottleneck for the Classification Model.

The loss for the joint objective becomes: $\mathcal{L}_{joint} = \mathcal{L}_{TE} + \lambda \mathcal{L}_{clf}$

$$\mathcal{L}_{TE} = \sum_k \sum_{j=1}^{m} -p_{k,j}^{true} \log p_{k,j} \qquad \mathcal{L}_{clf} = \sum_k \sum_{j=1}^{m} -c_{k,j}^{true} \log c_{k,j}$$

where **λ**: weight of the two-step classification objective,
m: total number of classes,
$\mathbf{p}^{\mathbf{true}}_{\mathbf{k,j}}$ and $\mathbf{c}^{\mathbf{true}}_{\mathbf{k,j}}$: binary label of sample k belonging to class j, and
$\mathbf{p}_{\mathbf{k,j}}$ and $\mathbf{c}_{\mathbf{k,j}}$: probability of predicted label for sample k to be class j.

Textual Entailment, Two-Way Classification



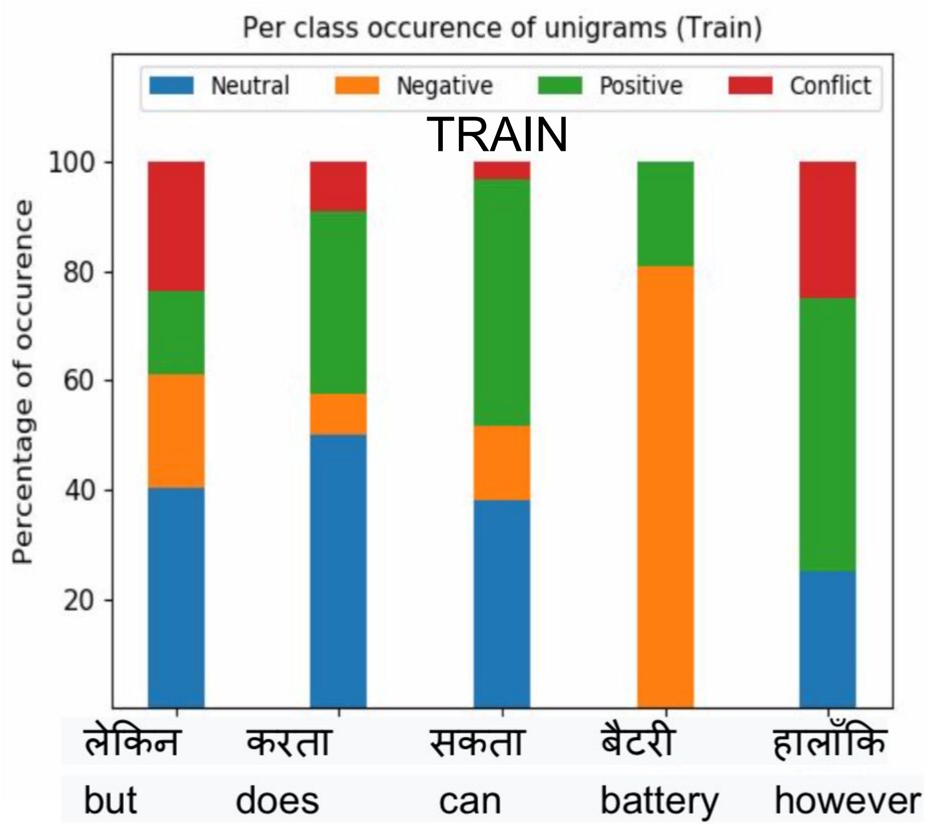
Original Dataset: Semantic Phenomenon

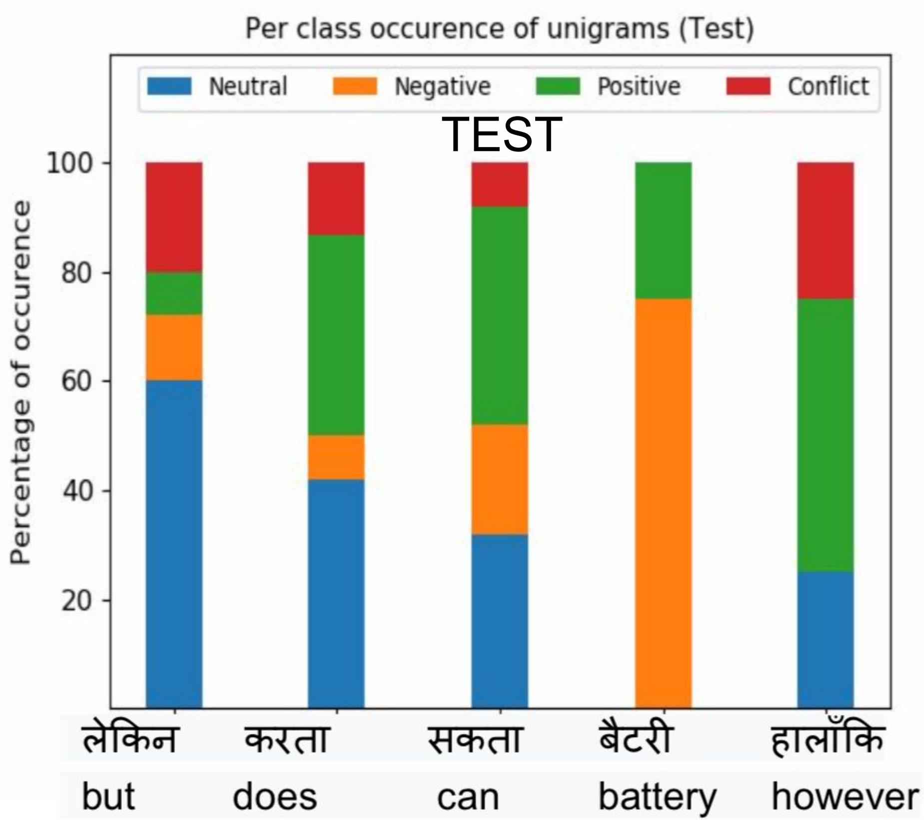
Semantic Phenomenon	Dataset Name	# Classes	# Data points	
			Classification	NLI (Recasted)
Sentiment Analysis	Product Review	4	5417	26000
Emotion Recognition	BHAAV	5	20304	105582
Discourse Analysis	Hindi Discourse	5	10472	54458
Topic Modelling	BBC News	6	4335	20752

	Datasets			
	PR	**BH**	**HDA**	**BBC**
Recasted TE / NLI Data				
# Classes	2	2	2	2
# Train	17336	64972	33508	15556
# Dev	4328	20300	10470	2592
# Test	4336	20310	10480	2604

	Datasets			
	PR	**BH**	**HDA**	**BBC**
Direct Classification Data				
# Classes	4	5	5	6
# Train	4334	16243	8377	3889
# Dev	541	2030	1047	216
# Test	542	2031	1048	217

ARTIFACT PROBLEM : MODEL LEARNS SPURIOUS PATTERN





Experiments And Results

The experiments are conducted so as to answer the following questions:

1. Representations effectiveness to derive **logical entailment** in context-hypothesis pairs on re-casted data?

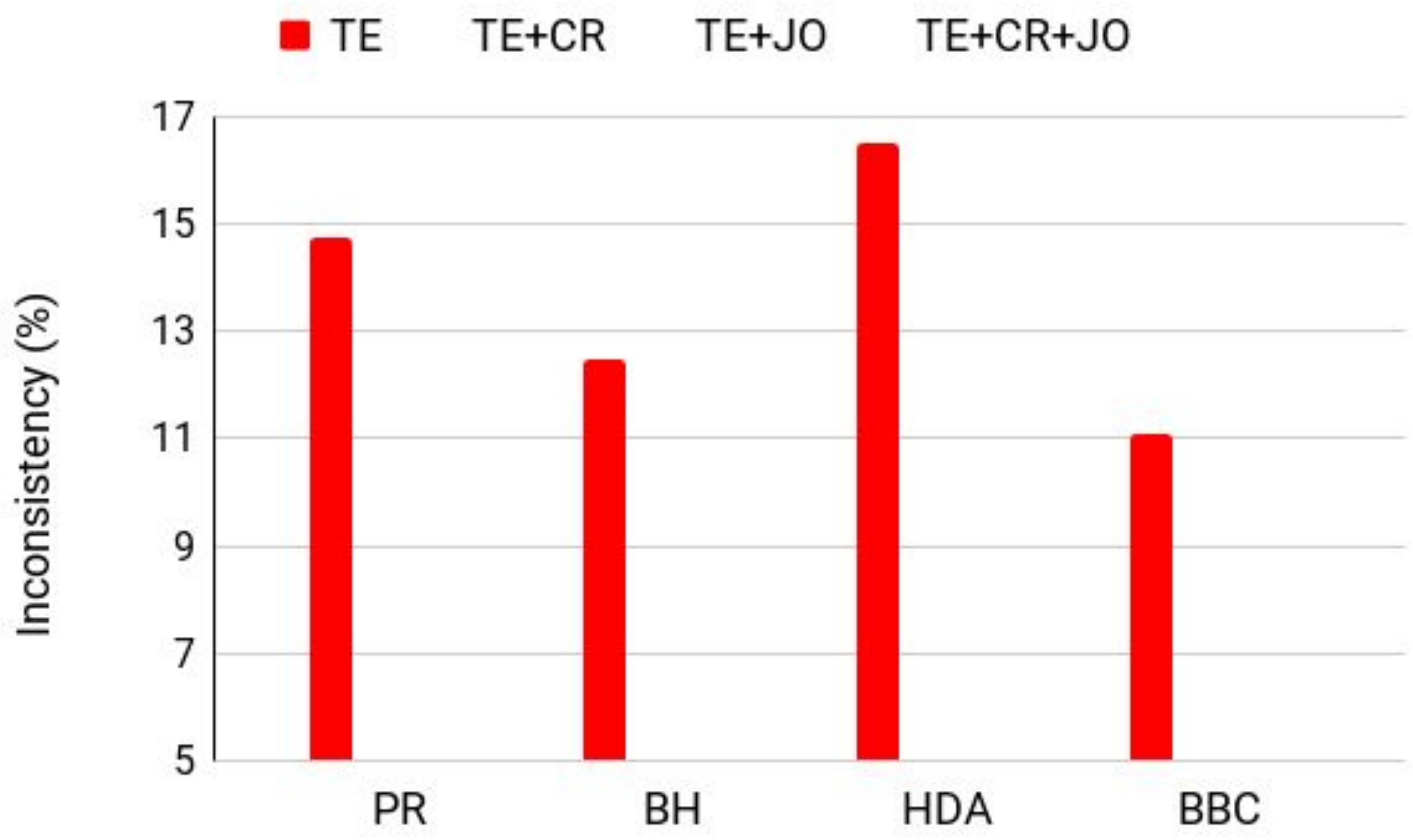
   How **consistent/inconsistent** are such representation with their own beliefs? Also, does **consistency regulariser** help in to **mitigate** such model inconsistency?

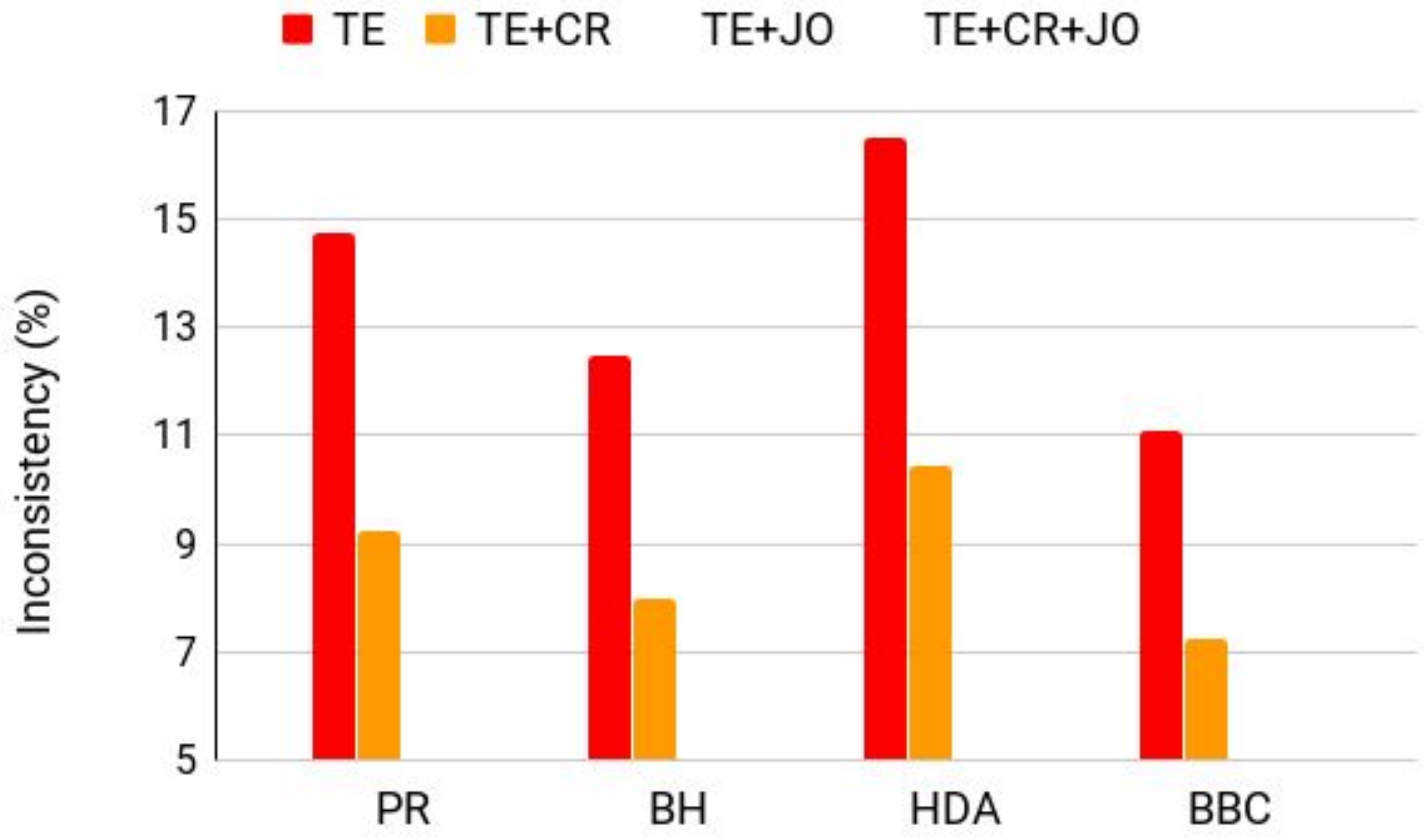
2. Do sentence representation models work well for classification task?

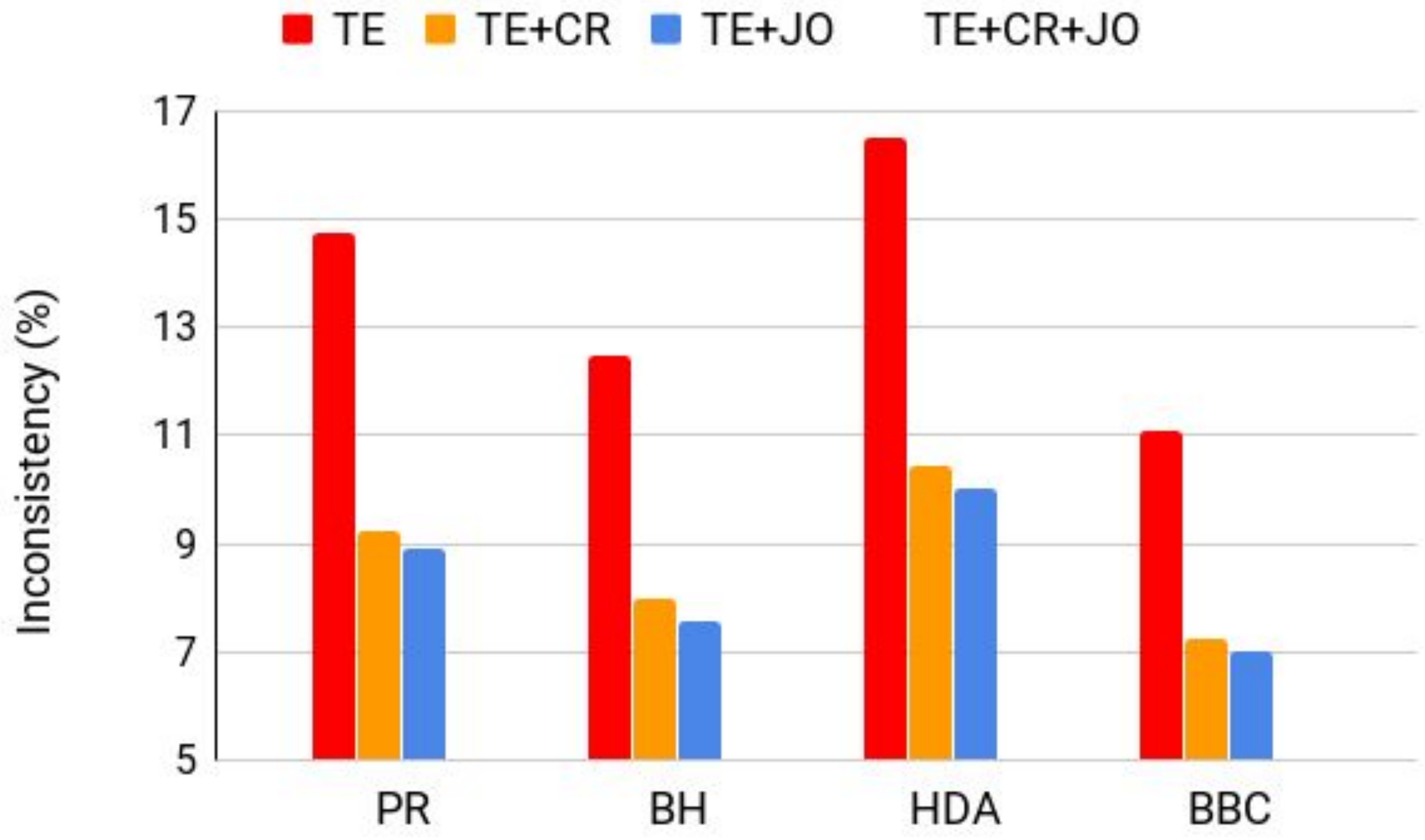
   Can re-casted NLI data be queried to **retrieve ground truth classification** label using two-step classification?

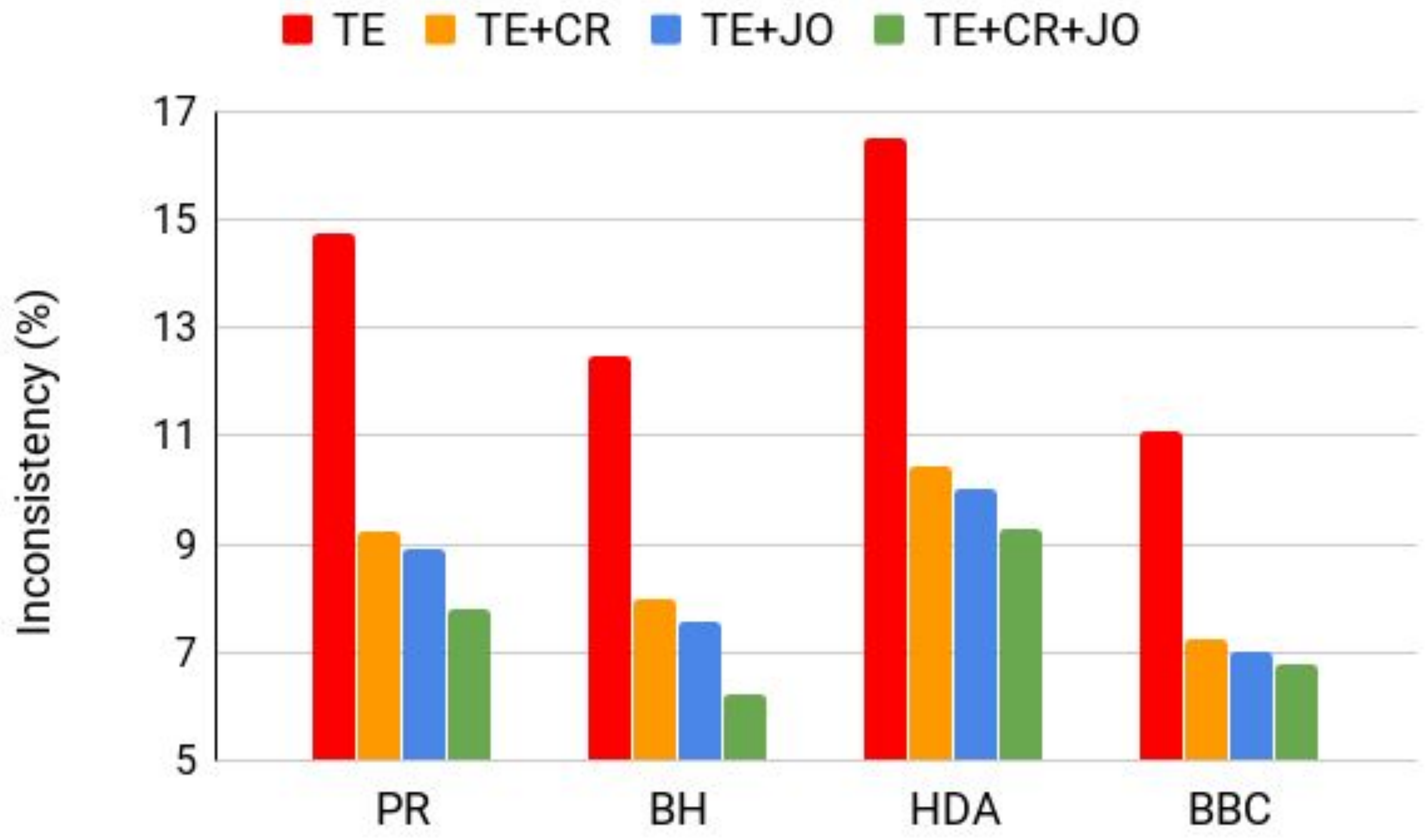
   Does our feedback through **joint training objective** further useful?

*All results are tested for statistically significant


TE
TE+CR
TE+JO
TE+CR+JO
Inconsistency (%)
17
15
13
11
9
7
5
PR
BH
HDA
BBC


TE
TE+CR
TE+JO
TE+CR+JO
Inconsistency (%)
17
15
13
11
9
7
5
PR
BH
HDA
BBC

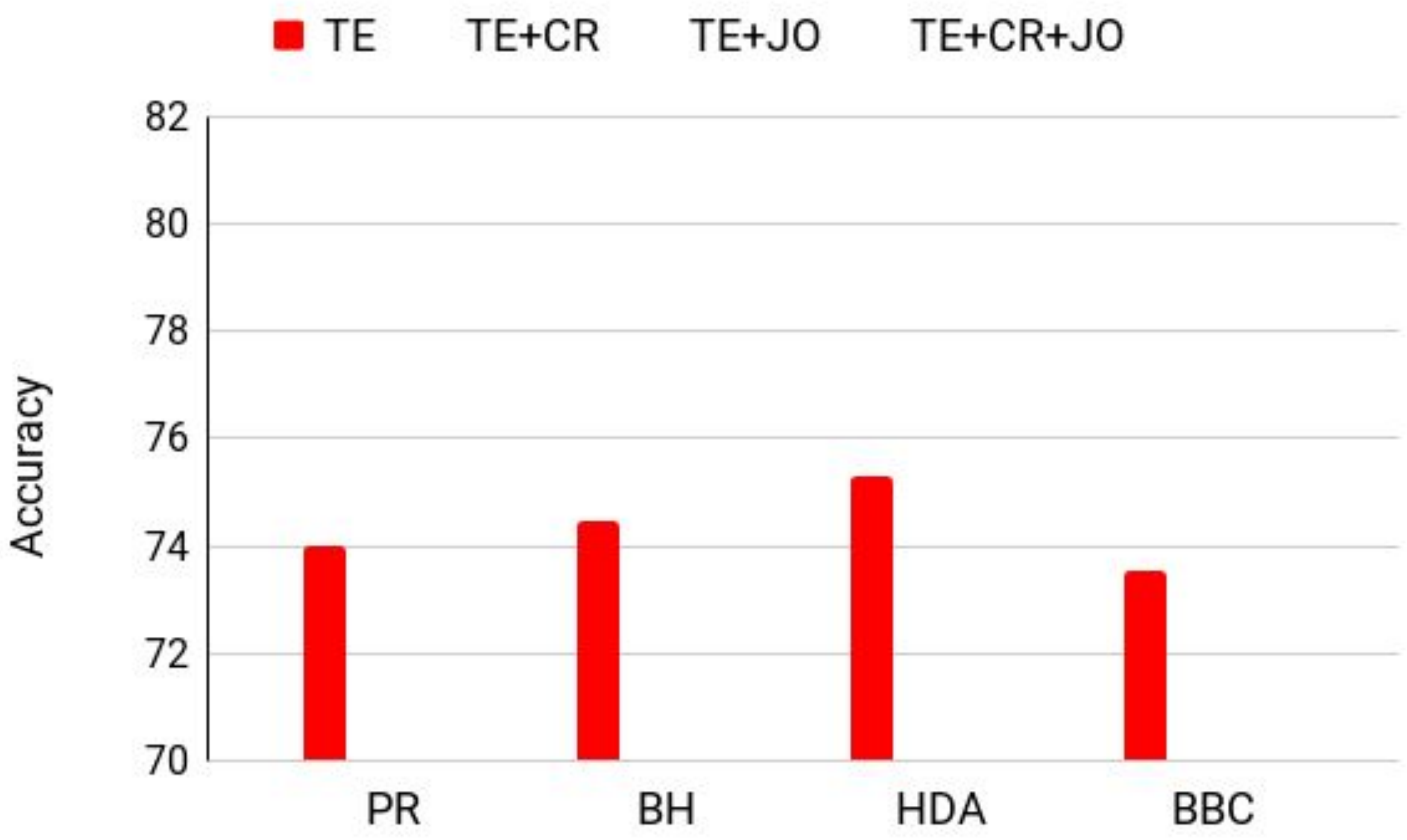

TE
TE+CR
TE+JO
TE+CR+JO
Inconsistency (%)
17
15
13
11
9
7
5
PR
BH
HDA
BBC

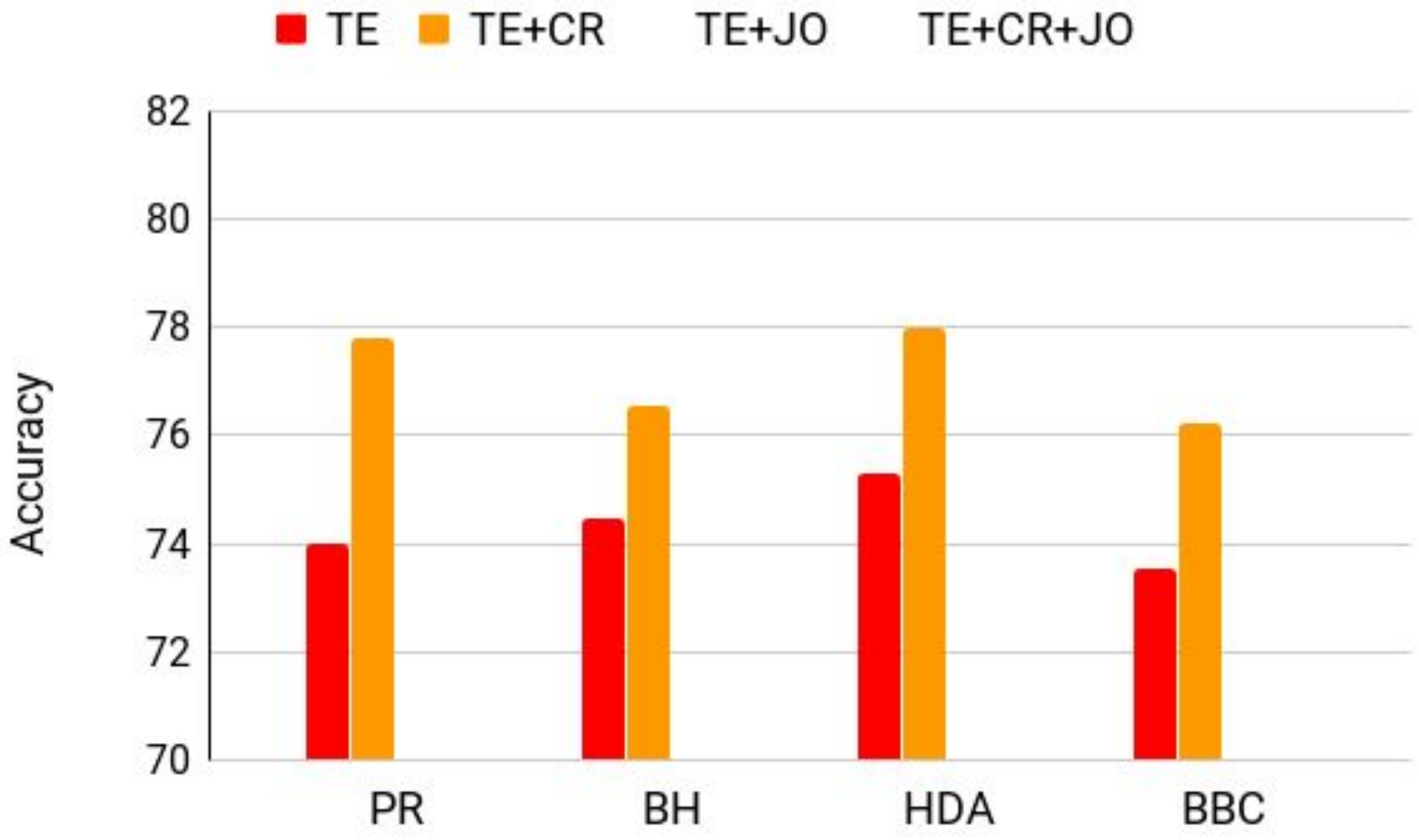


TE
TE+CR
TE+JO
TE+CR+JO
Inconsistency (%)
17
15
13
11
9
7
5
PR
BH
HDA
BBC

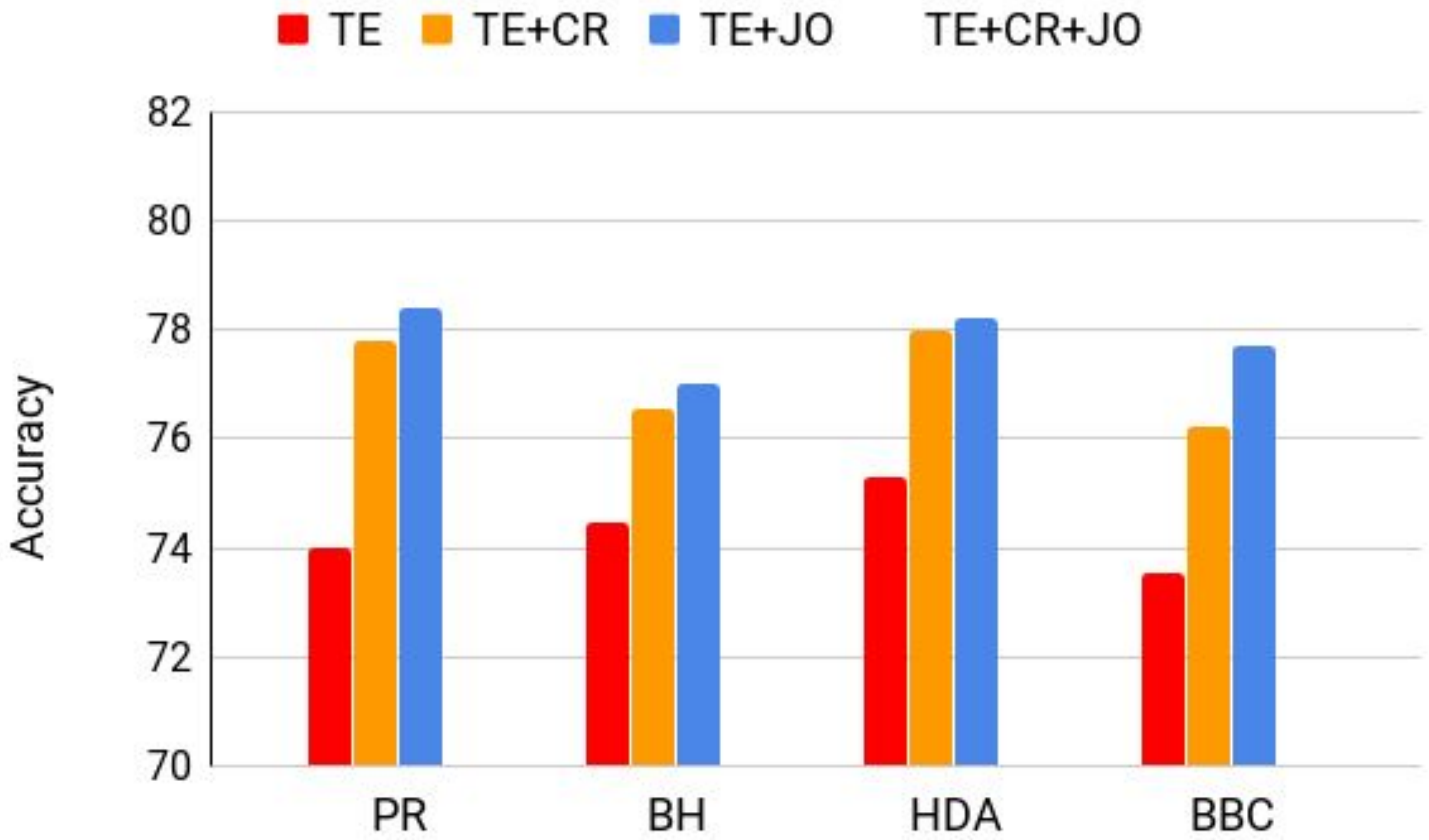
INCONSISTENCY CONCLUSION

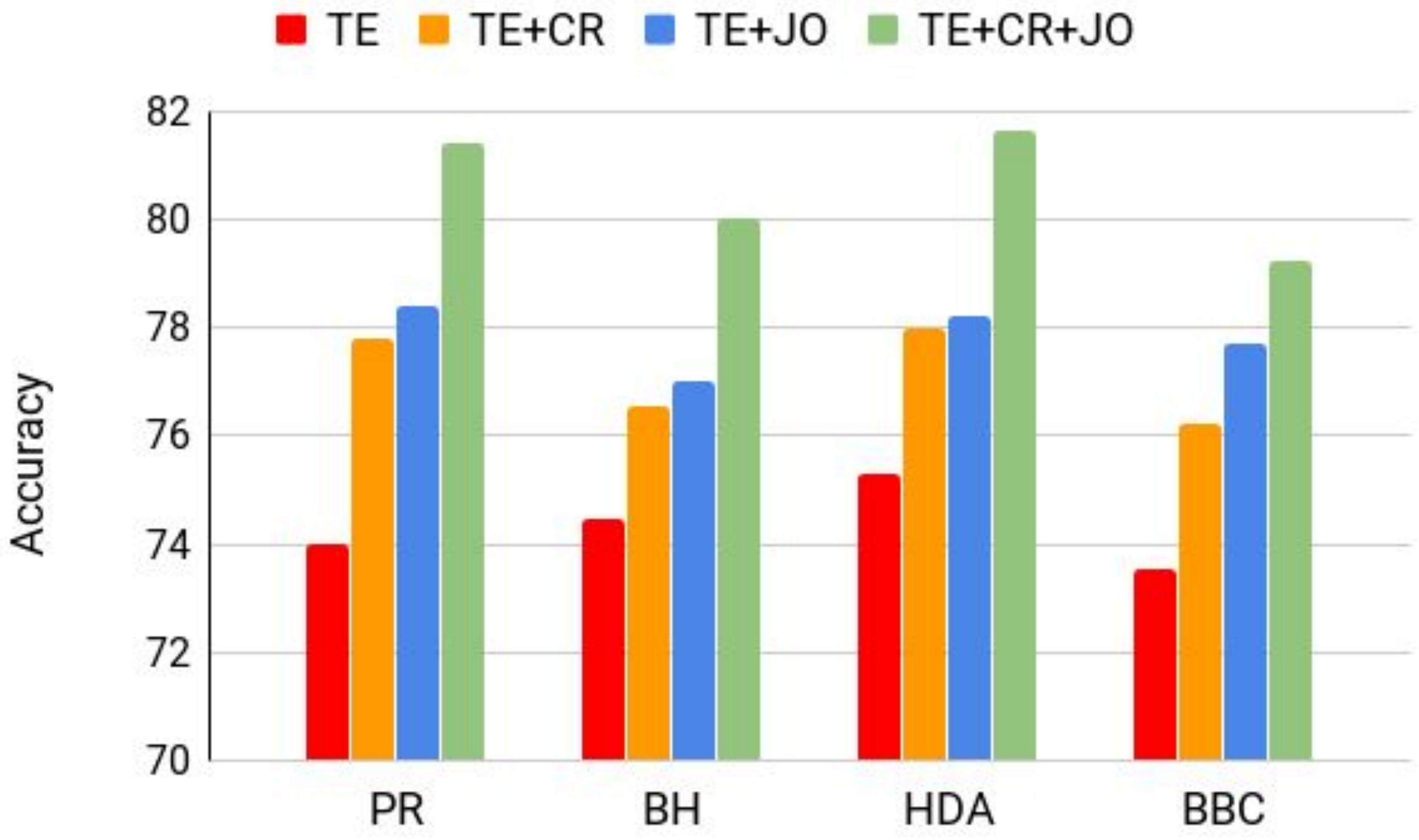
TEXTUAL ENTAILMENT PERFORMANCE


TE
TE+CR
TE+JO
TE+CR+JO
Accuracy
82
80
78
76
74
72
70
PR
BH
HDA
BBC



TE
TE+CR
TE+JO
TE+CR+JO
Accuracy
82
80
78
76
74
72
70
PR
BH
HDA
BBC

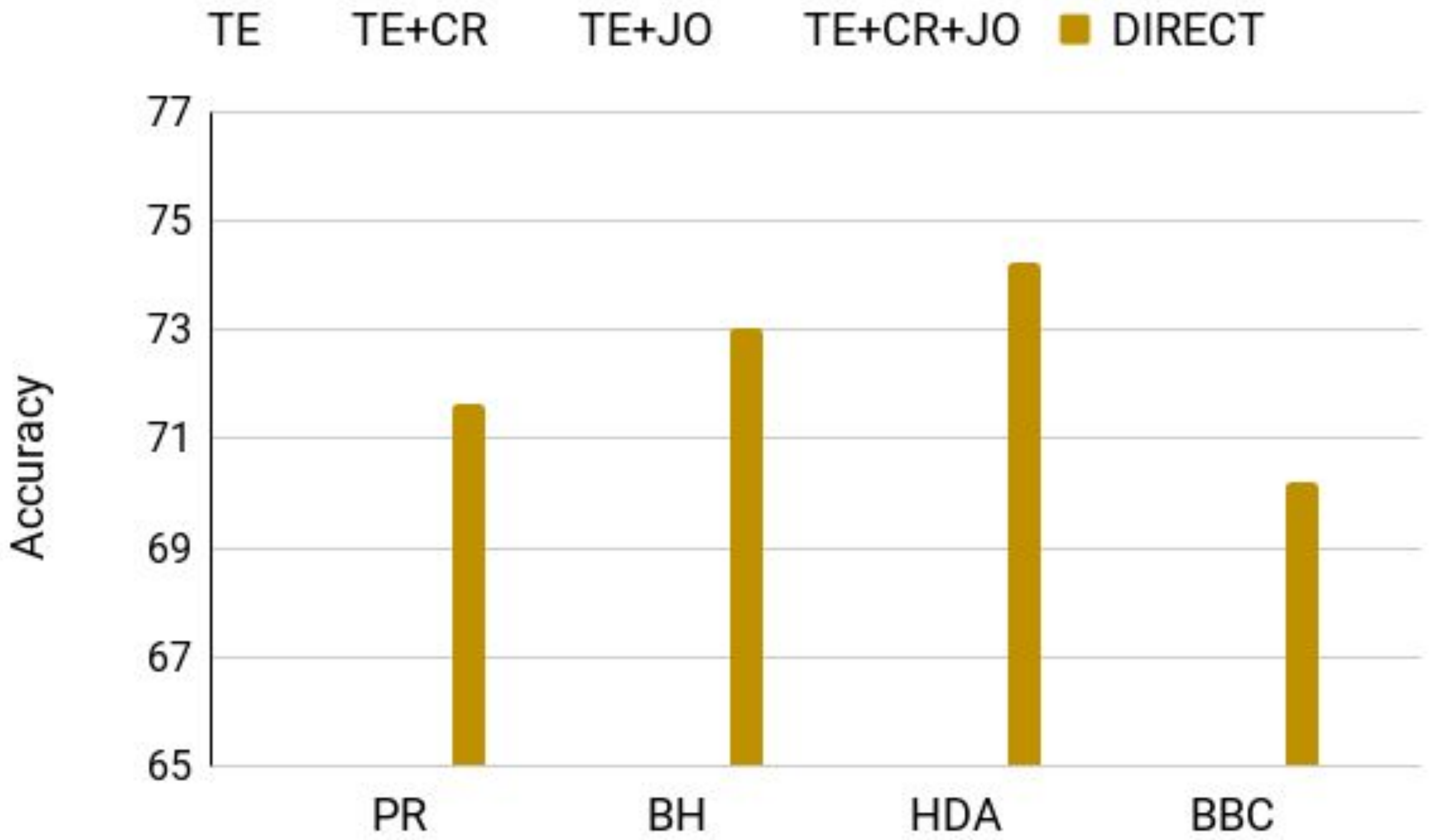

TE
TE+CR
TE+JO
TE+CR+JO
Accuracy
82
80
78
76
74
72
70
PR
BH
HDA
BBC

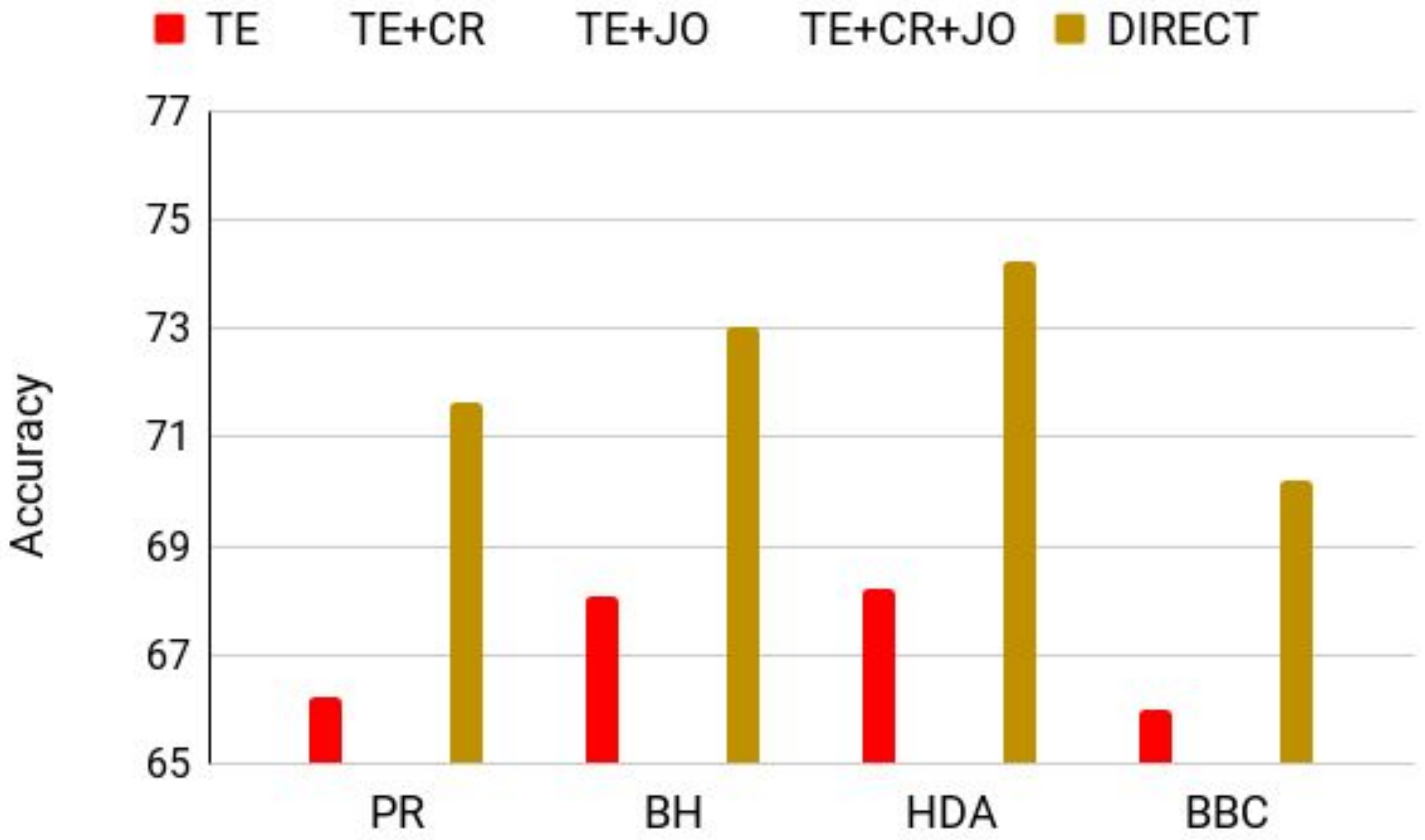


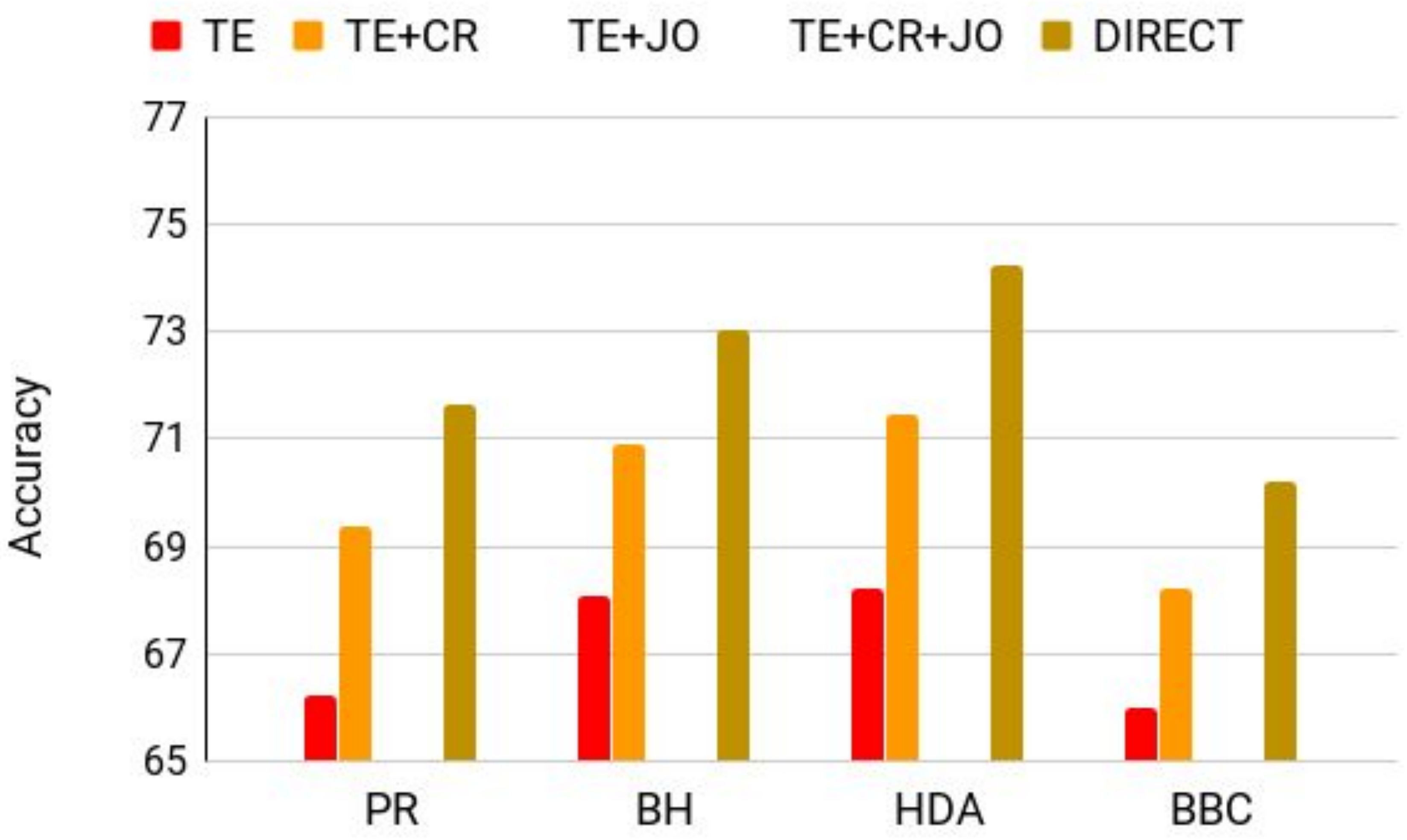
TE
TE+CR
TE+JO
TE+CR+JO
Accuracy
82
80
78
76
74
72
70
PR
BH
HDA
BBC

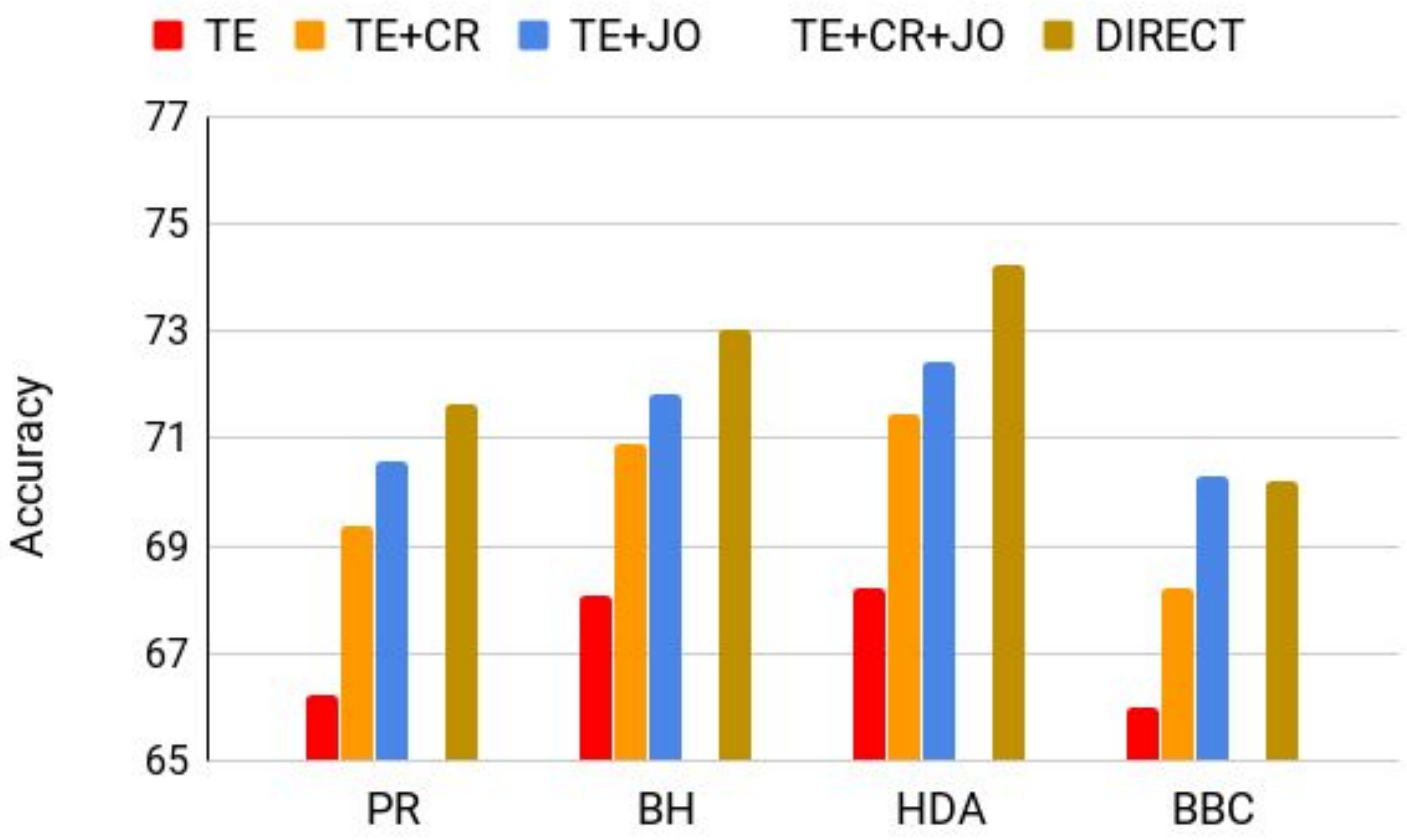
Textual Entailment Conclusion

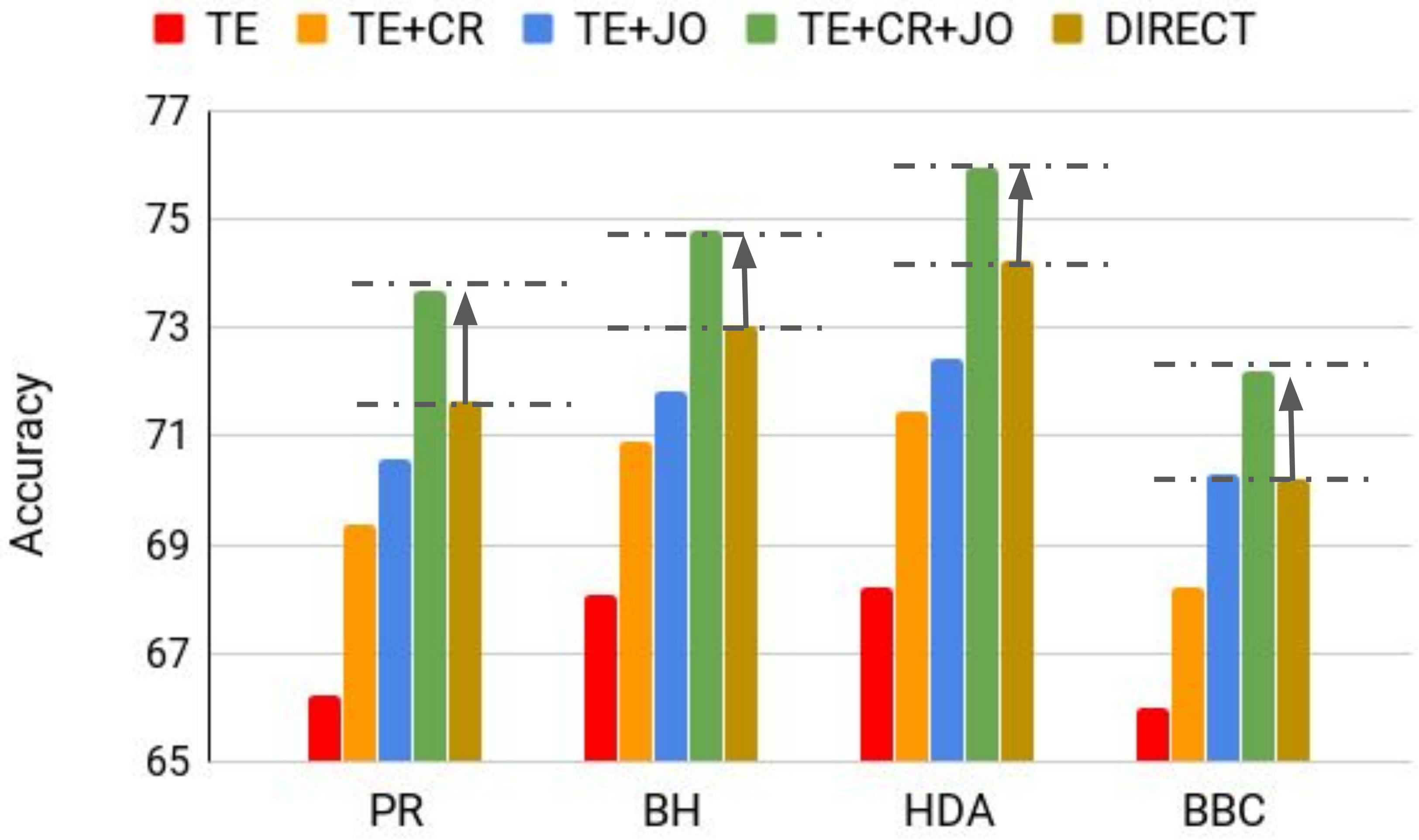
CLASSIFICATION PERFORMANCE


TE
TE+CR
TE+JO
TE+CR+JO
DIRECT
Accuracy
77
75
73
71
69
67
65
PR
BH
HDA
BBC


TE
TE+CR
TE+JO
TE+CR+JO
DIRECT
Accuracy
77
75
73
71
69
67
65
PR
BH
HDA
BBC


TE
TE+CR
TE+JO
TE+CR+JO
DIRECT
Accuracy
77
75
73
71
69
67
65
PR
BH
HDA
BBC


TE
TE+CR
TE+JO
TE+CR+JO
DIRECT
Accuracy
77
75
73
71
69
67
65
PR
BH
HDA
BBC


TE
TE+CR
TE+JO
TE+CR+JO
DIRECT
Accuracy
77
75
73
71
69
67
65
PR
BH
HDA
BBC

Take Away & Benefit Of Recasting

In-expensive dataset creation using automatic rule over template base set

Label balance re-casted data is robust to artifacts & spurious correlations

Structural regulariser over re-casted data can remove model inconsistency

Improve representation, performance, and interpretability on downstream tasks

Diverse semantic phenomena and multiple domain dataset can be unifiedly

References

[1] Product Review Dataset

[2] BHAAV Dataset

[3] Hindi Discourse Analysis Dataset

[4] BBC News Dataset

ORIGINAL DATASET: SEMANTIC PHENOMENON

1. Sentiment Analysis. Product Review Dataset (PR)} [1]. online user reviews for different products in ***Hindi,* 5417 sentences.** 4 sentiment classes: **Positive**, **Negative**, **Neutral** and **Conflict**.

2. Emotion Analysis. BHAAV Dataset (BH) [2]. This dataset comprises of **20,304 *Hindi* sentences** collected from 230 short stories ranging to diverse genres. It comprises of five emotion categories: **Joy**, **Anger**, **Suspense**, **Sad** and **Neutral**.

3. Discourse Analysis. Hindi Discourse Analysis Dataset (HDA) [3]. This dataset consists of **10,472 sentences** for analysing different **modes of discourse**. Classes comprises of **Argumentative**, **Descriptive**, **Dialogic**, **Informative** and **Narrative**.

4. Topic Modelling. Hindi BBC News Dataset (BBC) [4]. Comprises of **4,335 hindi news headlines**. In the original dataset, there are 14 classes, merging similar labels give 6 classes: **International**, **News**, **India**, **Sports**, **Science** and **Entertainment.**