



15th International Natural Language Generation
Conference

HinglishEval Generation Challenge on Quality Estimation of Synthetic Code-Mixed Text: Overview and Results

Vivek Srivastava
Researcher
TCS Research,
India

Mayank Singh
Assistant Professor
IIT Gandhinagar,
India

tcs Research



Motivation

- Mixing words and phrases from two (or more) languages.
- Code-mixed text is “*noisy*” in nature.
- Traditional NLG evaluation strategies fail.
- How do we measure the quality of the code-mixed text?
 - Generated synthetically (typical output from an NLG system)
 - Mix of Hindi-English languages
- Do humans have a contradictory notion of the quality of the code-mixed text?

The HinGE dataset

Parallel
Hindi-English
sentences

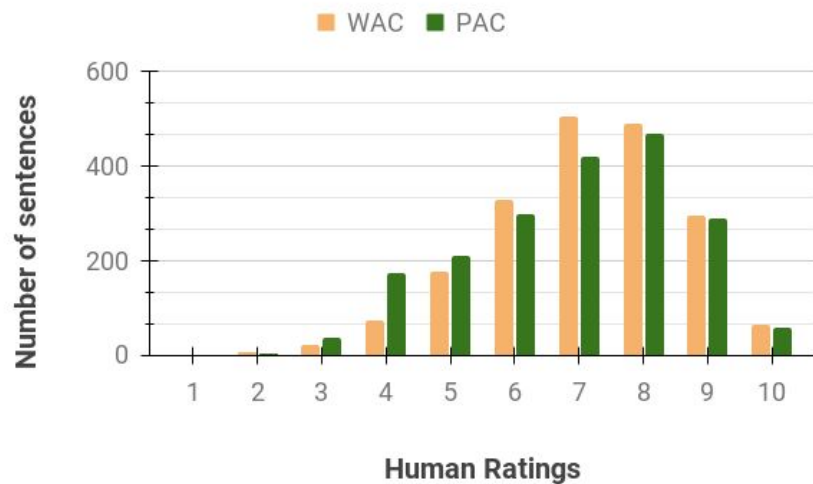
Machine-generated
Hinglish sentences

English	Hindi	Human-generated Hinglish	WAC	PAC
The reward of goodness shall be nothing but goodness.	अच्छाई का बदला अच्छाई के सिवा और क्या हो सकता है?	The reward of achai shall be nothing but achai.	reward ka badla reward ke nothing aur kya ho sakta hai Rating1: 7 Rating2: 4	reward of goodness goodness ke siva aur kya ho sakta hai Rating1: 9 Rating2: 7
		Goodness ka badla goodness ke siva aur kya ho sakta hai.		
		Achai ka badla shall be nothing but achai.		

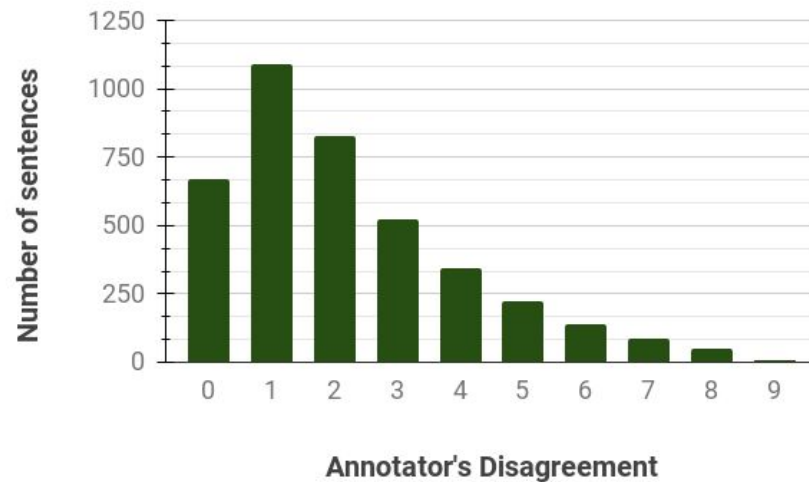
Multiple
reference
Hinglish
sentences

2 human-ratings
to generated
Hinglish
sentences

The two sub tasks

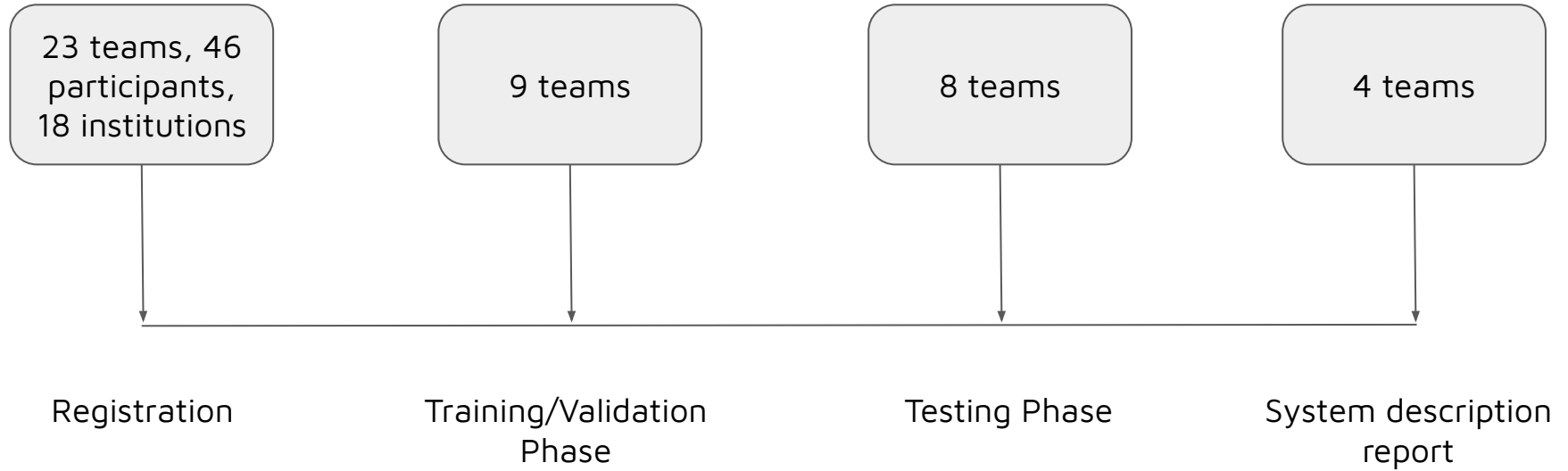


Subtask 1



Subtask 2

Participation overview



Experiments

<i>Team Name</i>	<i>Embedding model</i>	<i>Additional features</i>	<i>Prediction model</i>
Baseline	M-BERT	×	Fine-tuned M-Bert with one hidden layer neural network
IIITH	XLM-R and LABSE	CMI, Number of switch points, and burstiness	MLP regression
MU	BERT ^[1] and LABSE	×	Catboost classifier
BITS	M-BERT	×	Fully connected two layer neural network
JU	GloVe and one-hot encoding	×	LSTM

[1] <https://huggingface.co/niksss/Hinglish-HATEBERT>

Results

Team Name	Subtask 1			Subtask 2	
	FS	CK	MSE	FS	MSE
Baseline	0.26637 (1)	0.09922 (1)	2.00000 (1)	0.14323 (8)	5.00000 (3)
IITH	0.25734 (2)	0.09858 (2)	2.00000 (1)	0.23523 (3)	3.00000 (1)
MU	0.25062 (3)	0.08153 (3)	2.00000 (1)	0.26115 (1)	3.00000 (1)
BITS	0.21796 (5)	0.07337 (5)	3.00000 (2)	0.23940 (2)	4.00000 (2)
JU	0.11582 (9)	0.00337 (8)	6.00000 (3)	0.18331 (6)	5.00000 (3)

Results on the test set. Number inside a bracket represent relative rank in respective task for a particular metric.

Open challenges

- New metric(s) for code-mixed NLG evaluation.
 - Checklist for quality evaluation of code-mixed text.
- Generalizability with other code-mixed languages
- Releasing the dataset

Questions | Feedback | Comments



Email

srivastava.vivek2@tcs.com

Webpage

sites.google.com/view/vivek-srivastava

Twitter

[@srivivek36](https://twitter.com/srivivek36)



singh.mayank@iitgn.ac.in

[mayank4490.github.io](https://github.com/mayank4490)

[@mayank_iitgn](https://twitter.com/mayank_iitgn)