TEXT STYLE TRANSFER FOR PARAPHRASING

INFORMATION RETRIEVAL AND EXTRACTION

PROJECT

TEAM – 07

PRANSHU NEMA
ADVAIT SHRIVASTAVA
AYUSH LAKSHAKAR

AGENDA

Problem Description

Dataset

Implementation

Evaluation

PROBLEM DESCRIPTION

Text Style Transfer

The task of style transfer on text data involves changing the style of a given sentence while preserving its semantics.

The goal is to automatically control the style attributes of text without any loss in the content.

TST has many immediate applications.

One such application is intelligent bots, for which users prefer distinct and consistent persona (e.g., empathetic) instead of emotionless or inconsistent persona.

Another application is the development of intelligent writing assistants

(input, target style) → output

ensure content of input and output are similar

O, wilt thou leave me so unsatisfied? → oh u will leave me so sad?? :(

(Shakespeare) (Twitter)

Source sentence x:	*"Come and sit!"*	Source attribute a:	*Informal*
Target sentence x':	*"Please consider taking a seat."*	Target attribute a':	*Formal*

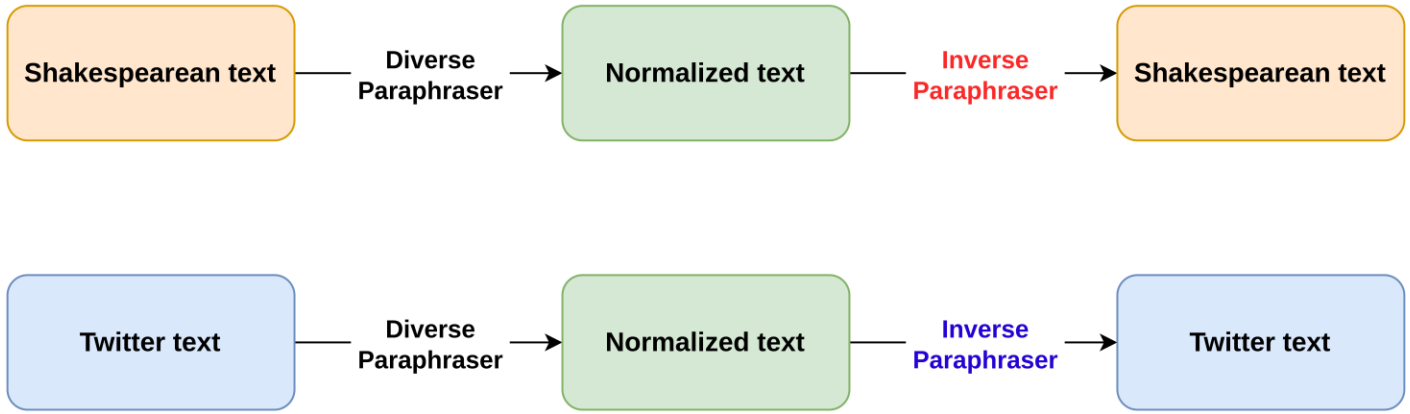
DATASET

- Shakespeare author imitation dataset
- Formality transfer dataset
- Corpus of Diverse Styles dataset (CDS)
- Filtered ParaNMT - 50M

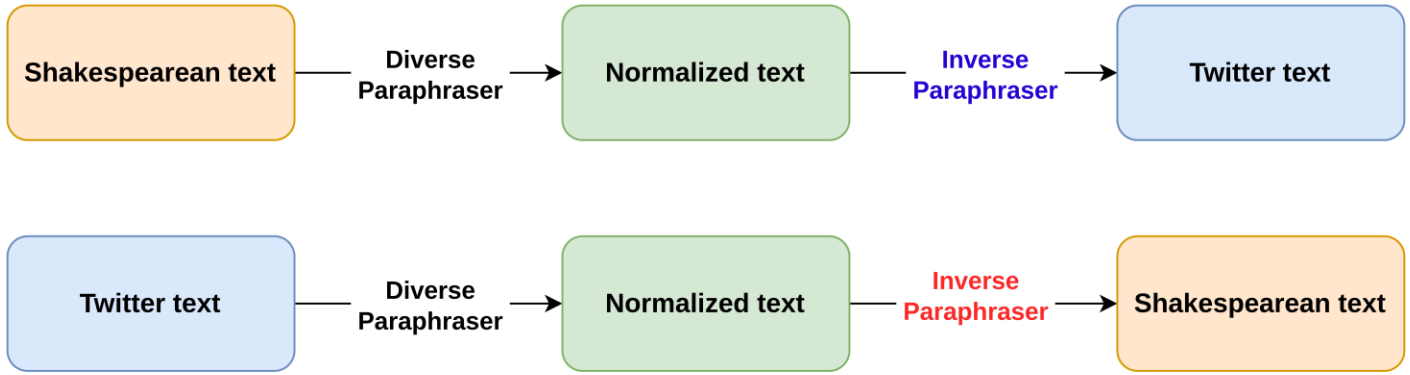
IMPLEMENTATION

Proposed Architecture

Training



Inference



Diverse Paraphraser

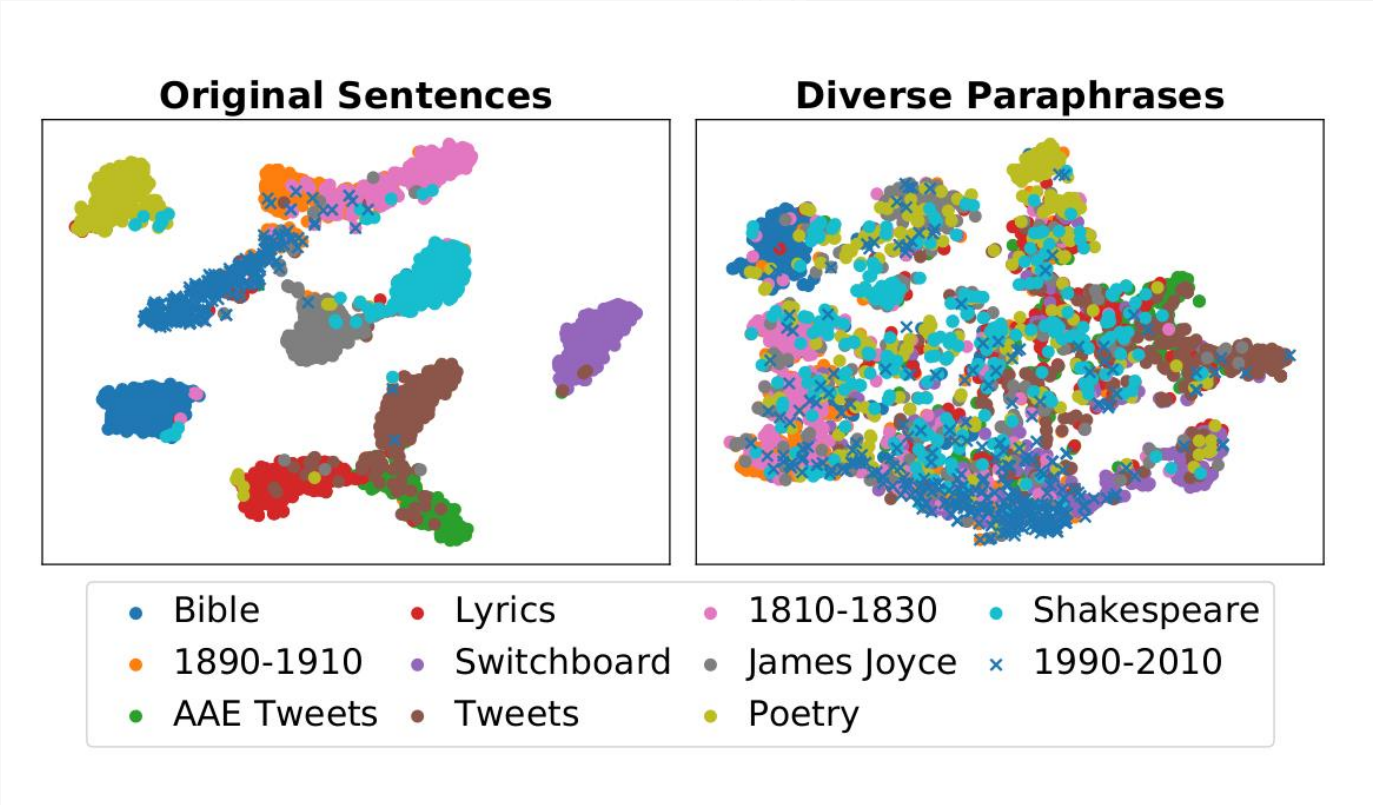


Figure - Diverse paraphrasing normalizes sentences by removing stylistic identifiers.

EVALUATION

The commonly used practice of evaluation considers the following three criteria :-

Transferred style strength: Common way of measuring transfer success is to train a classifier to identify the style of a transferred sentence and report its accuracy ACC on generated sentences.

Semantic preservation: Measuring how much a transferred sentence deviates in meaning from the input. Measuring semantic similarity using the subword embedding-based SIM model of Wieting et al. (2019), which performs well on semantic textual similarity (STS).

Fluency: Language model perplexity, Train a seperate LL classifier on corpus like coLA.

THANKYOU