TRANSFORMER MODEL PERFORMANCE ON AUTOMATIC SPEECH RECOGNITION OF INDIAN NEWS BULLETINS IN 30 LANGUAGES

CHARLES REDMON

ABSTRACT. Four Transformer-based automatic speech recognition (ASR) models – Whisper, AI4Bharat, ... – were evaluated against a corpus of news bulletins from All India Radio spanning 30 languages and 60 locations.

1. INTRODUCTION

The past five years have seen a rapid rise in ASR model performance with relatively minimal feature engineering as a consequence of the Transformer model [ ] and wav2vec embeddings from speech [ ], among other architectural advances [other key citations here].

1.1. Model summaries.

1.1.1. *Whisper.*

There now is your insular city of the Manhattoes, belted round by wharves as Indian isles by coral reefs - commerce surrounds it with her surf. Right and left, the streets take you waterward. Its extreme down-town is the battery, where that noble mole is washed by waves, and cooled by breezes, which a few hours previous were out of sight of land. Look at the crowds of water-gazers there.

Anyone caught using formulas such as $\sqrt{x+y} = \sqrt{x} + \sqrt{y}$ or $\frac{1}{x+y} = \frac{1}{x} + \frac{1}{y}$ will fail.

The binomial theorem is

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

A favorite sum of most mathematicians is

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

Likewise a popular integral is

$$\int_{-\infty}^{\infty} e^{-x^2}\,\mathrm{d}x = \sqrt{\pi}$$

Theorem 1.1.1.1. *The square of any real number is non-negative.*

Proof. Any real number x satisfies $x > 0$, $x = 0$, or $x < 0$. If $x = 0$, then $x^2 = 0 \geq 0$. If $x > 0$ then as a positive time a positive is positive we have $x^2 = xx > 0$. If $x < 0$ then $-x > 0$ and so by what we have just done $x^2 = (-x)^2 > 0$. So in all cases $x^2 \geq 0$. □

2. INTRODUCTION

This is a new section. You can use tables like.

2.1. Things that need to be done.

Prove theorems, such as Theorem 2.1.1.

Theorem 2.1.1. *The Riemann hypothesis is true.*

Proof. This is left as an exercise to the reader, given the complexity of the theorem. □

DEPARTMENT OF LANGUAGE AND LINGUISTICS, UNIVERSITY OF ESSEX
Email address: c.redmon@essex.ac.uk
URL: lab.speechsystems.org/usr/redmon