**BERTologiCoMix: How does Code-Mixing interact with Multilingual BERT?**

Sebastin Santy, Anirudh Srinivasan, Monojit Choudhury
Microsoft Research, India

---

**Code-Mixing + BERTology = BERTologiCoMix**

**Code Mixing and Code-Switching**

- Life ko face kiijiye with himmat and faith in yourself
  "Face life with courage and faith in self"

- She lives en una casa blanca
  "She lives in a white house"

**BERTology**

- Series of studies probing BERT and its representations (Rogers et. al., 2020)

Questions we ask:
- What type of CM is ideal for mBERT finetuning?
- What changes happen to mBERT while finetuning?

**Types of Code-Mixing**

- \((l - CM)\) – lexical Code-Mixing (random replacement)
- \((g - CM)\) – generated Code-Mixing (synthetic) (Pratapa et. al., 2018)
- \((r - CM)\) – real Code-Mixing (naturally occurring)

**Downstream Task Experiments**

<table>
<thead>
<tr>
<th>model</th>
<th>enes</th>
<th>enhi</th>
<th>enes</th>
<th>enhi</th>
<th>enhi</th>
<th>enhi</th>
<th>enes</th>
<th>enhi</th>
<th>QA</th>
<th>NLI</th>
</tr>
</thead>
<tbody>
<tr>
<td>m()</td>
<td>67.01±0.15</td>
<td>58.42±0.15</td>
<td>59.50±0.09</td>
<td>75.55±0.06</td>
<td>93.35±0.02</td>
<td>87.49±0.02</td>
<td>63.40±0.05</td>
<td>95.99±0.08</td>
<td>95.80±0.05</td>
<td>71.95±0.08</td>
</tr>
<tr>
<td>m(l,CM)</td>
<td>m(g,CM)</td>
<td>m(r,CM)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

GLUECoS Benchmark (Khanuja et al., 2020) consists of varied code-mixing tasks

Sentiment, NER, POS, Language ID, QA, NLI | English-Spanish \((enes)\) and English-Hindi \((enhi)\)

**Responsivity to Code-Mixing**

Build a classifier to distinguish between Monolingual and Code-Mixed sentences using BERT attention head representations by measuring responsivity \((R_{x,y})\) (analogous to calculating information gain of features)

**References**