**LARD: Large-scale Artificial Disfluency Generation**

**T. Passali, T. Mavropoulos, G. Tsoumakas, G. Meditskos, S. Vrochidis**

---

### Introduction

- Virtual assistants and spoken dialogue systems are increasingly used in many applications.
- **Disfluencies**: Interruptions, self-corrections, false-starts, repetitions etc.
- **Disfluency detection**: Detection of disfluent regions in spoken language transcripts

### Motivation

- Existing datasets do not contain sufficiently all the different types of disfluencies.
- **Example**: Switchboard [1] contains only 40K/160K utterances with more than 50% of repetitions (most trivial class).
- Existing augmentation techniques use simplistic rules and are not capable of generating all different kinds of disfluencies.

---

**Generating artificial disfluencies from fluent text**

- **Replacement Algorithm**
  - Given a fluent sequence:
  - 1. Randomly extract a repair candidate (noun, verb, adjective)
  - 2. Generate synonyms and antonyms for the selected candidate
  - 3. Replace with a synonym or antonym with/without a repair cue

- **Restart Algorithm**
  - Given two or more fluent sequences:
  - 1. Randomly pick two sequences
  - 2. Split the first sequence in a random position
  - 3. Combine the broken sequence with the second unbroken one

---

**LARD Dataset**

<table>
<thead>
<tr>
<th>Dataset Statistics</th>
<th># repetitions</th>
<th>23398</th>
<th># replacements</th>
<th>23398</th>
<th># restarts</th>
<th>23398</th>
<th># fluencies</th>
<th>23398</th>
<th># total</th>
<th>95992</th>
</tr>
</thead>
</table>

A new large-scale artificial and balanced dataset for disfluency detection based on an existing fluent dataset: Schema-Guided Dataset [2].

---

**Results**

<table>
<thead>
<tr>
<th></th>
<th>Prec</th>
<th>Rec</th>
<th>FL</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detection</td>
<td>97.63</td>
<td>97.61</td>
<td>97.62</td>
<td>-</td>
</tr>
<tr>
<td>Classification</td>
<td>97.31</td>
<td>97.30</td>
<td>97.29</td>
<td>-</td>
</tr>
<tr>
<td>Extraction</td>
<td>98.12</td>
<td>96.60</td>
<td>97.30</td>
<td>-</td>
</tr>
<tr>
<td>Correction</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>86.48</td>
</tr>
</tbody>
</table>

**Table 2: Experimental results on LARD dataset**

<table>
<thead>
<tr>
<th></th>
<th>Switchboard (detection)</th>
<th>LARD (detection)</th>
<th>LARD (classification)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repetitions</td>
<td>85.42</td>
<td>99.57</td>
<td>99.5</td>
</tr>
<tr>
<td>Replacements</td>
<td>54.52</td>
<td>99.67</td>
<td>98.39</td>
</tr>
<tr>
<td>Restarts</td>
<td>19.6</td>
<td>95.08</td>
<td>93.89</td>
</tr>
</tbody>
</table>

**Table 3: Accuracy (%) for different disfluency classes (repetitions, replacements and restarts) and models trained on different datasets.**

---

**References**


---

**Acknowledgements**

This work is partially funded by the European Commission as part of its H2020 Programme, under the contract number 870930-IA WELCOME Project.

[welcome-h2020.eu]