Neural metaphor detection for Slovene

Matej Klemen
(supervisor: prof. Marko Robnik Šikonja)
The problem

- **what** is a metaphor, **why** detect them

His words *cut* deeper than a knife.

- the tools:

  - CroSloEngual BERT 1.1
    - Slovenian RoBERTa contextual embeddings model: SloBERTa 2.0
    - **~110M parameters**
  - Metaphor corpus KOMET 1.0
    - **~250K tokens, ~5% metaphors**
  - Corpus of metaphorical expressions in spoken Slovene language G-KOMET 1.0
    - **~50K tokens, ~2% metaphors**
## Results

- **F1 score**: correctness and coverage of the found metaphors

<table>
<thead>
<tr>
<th>Model</th>
<th>KOMET</th>
<th>G-KOMET</th>
</tr>
</thead>
<tbody>
<tr>
<td>(random)</td>
<td>0.114</td>
<td>0.016</td>
</tr>
<tr>
<td>SloBERTa</td>
<td>0.597</td>
<td>0.374</td>
</tr>
<tr>
<td>CSEBERT</td>
<td>0.552</td>
<td>0.310</td>
</tr>
<tr>
<td>(random)</td>
<td>0.035</td>
<td>0.018</td>
</tr>
<tr>
<td>SloBERTa</td>
<td>0.391</td>
<td>0.407</td>
</tr>
<tr>
<td>CSEBERT</td>
<td>0.265</td>
<td>0.368</td>
</tr>
</tbody>
</table>

*full dataset*

*noun + verb dataset*
Future work

- considering metaphors at token level is limited, consider spans
- cross-lingual transfer: does learning metaphor detection on other languages improve detection on Slovene?
- applications of automatic metaphor detection