SLOmet – Slovenian Commonsense Description

Adrian Mladenić Grobelnik, Erik Novak, Dunja Mladenic, Marko Grobelnik
Jožef Stefan Institute, Ljubljana, Slovenia
Outline

• Motivation
• Introduction
• Data Description
• Problem Description and Experimental Setting
• Experimental Results
• Demo
• Discussion
Motivation

• Commonsense is integral to our lives
  • Learn from mistakes & understand environment
  • Ex: Adrian drinks water, car

• Developed over time

• AI good at specialized tasks, unable to:
  • Infer commonsense from objects
  • Understand simple stories
Introduction

- AI for English commonsense
  - Deep Learning model trained on labelled sentences and entities
  - COMET (9 types), COMET-2020 (23 types)

- Multilingual commonsense
  - MultiCOMET - multilingually expand COMET
  - Proof of concept on Slovenian – English model, simple evaluation

- SLOmet
  - Translating commonsense data to Slovenian
  - Finetuning Slovenian language model
  - Comprehensive evaluation
Data Description

- Slovenian ATOMIC-2020 Dataset
  - 1.33 million commonsense descriptions
- Amazon Turk, open-text
- sentences & entities
- 23 commonsense descriptors
- Machine translation to Slovenian
  - 40,000 sentences and entities
  - 600,000 commonsense descriptions
Problem Description and Experimental Setting

• Given Slovene sentence/entity, predict Slovenian commonsense descriptions
• Finetuned two Slovenian GPT-2 language models
  • 1 million commonsense descriptions
  • macedonizer/sl-gpt2, gpt-janez
• Evaluation
  • Reference: English COMET-2020 GPT2-XL model
  • Metrics: BLEU, CIDEr, METEOR, ROUGE-L
  • 150,000 commonsense descriptions
Experimental Results: Metrics

- Comparable performance to English
  - English model trained for longer, is larger
  - Translation errors

- All struggled with 4-grams (BLEU-4)
- COMET gpt-janez outperforms COMET sl-gpt2

<table>
<thead>
<tr>
<th>Model</th>
<th>Language</th>
<th>BLEU-1</th>
<th>BLEU-2</th>
<th>BLEU-3</th>
<th>BLEU-4</th>
<th>CIDEr</th>
<th>METEOR</th>
<th>ROUGE-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMET sl-gpt2</td>
<td>Slovenian</td>
<td>0.297</td>
<td>0.150</td>
<td>0.086</td>
<td>0.058</td>
<td>0.487</td>
<td>0.207</td>
<td>0.383</td>
</tr>
<tr>
<td>COMET gpt-janez</td>
<td>Slovenian</td>
<td>0.324</td>
<td>0.174</td>
<td>0.108</td>
<td>0.076</td>
<td>0.508</td>
<td>0.225</td>
<td>0.397</td>
</tr>
<tr>
<td>COMET (GPT2-XL)</td>
<td>English</td>
<td>0.407</td>
<td>0.248</td>
<td>0.171</td>
<td>0.124</td>
<td>0.653</td>
<td>0.292</td>
<td>0.485</td>
</tr>
</tbody>
</table>

*description quality between 0 and 1*
Experimental Results: Examples

Marko je šel v trgovino (Marko went to the shop)

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>COMET sl-gpt2</th>
<th>COMET gpt-janez</th>
<th>COMET (GPT2-XL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>oWant</td>
<td>Nič</td>
<td>Nič</td>
<td>None</td>
</tr>
<tr>
<td>se zahvaliti osebiX</td>
<td>Nič</td>
<td>To give him a receipt</td>
<td></td>
</tr>
<tr>
<td>se zahvaliti</td>
<td>Nič</td>
<td>To give him a discount</td>
<td></td>
</tr>
<tr>
<td>IsBefore</td>
<td>Zaslužiti denar</td>
<td>Kupiti nekaj</td>
<td>PersonX buys a new car</td>
</tr>
<tr>
<td>V trgovino za hišne ljubljenčke</td>
<td>Kupiti nekaj</td>
<td>PersonX takes the car back home</td>
<td></td>
</tr>
<tr>
<td>V trgovino z živili</td>
<td>Kupiti nekaj</td>
<td>PersonX buys a new one</td>
<td></td>
</tr>
</tbody>
</table>

Avto (car)

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>COMET sl-gpt2</th>
<th>COMET gpt-janez</th>
<th>COMET (GPT2-XL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ObjectUse</td>
<td>Vožnja do trgovine</td>
<td>Priti do hiše</td>
<td>Drive to the store</td>
</tr>
<tr>
<td></td>
<td>Vožnja do hiše</td>
<td>Priti do hiše</td>
<td>Get to the store</td>
</tr>
<tr>
<td></td>
<td>Vožnja do cilja</td>
<td>Priti do hiše</td>
<td>Drive to the restaurant</td>
</tr>
<tr>
<td>HasProperty</td>
<td>Noro</td>
<td>Najden v avtomobilu</td>
<td>Found in parking lot</td>
</tr>
<tr>
<td></td>
<td>Vrata</td>
<td>Najden v avtomobilu</td>
<td>Found on road</td>
</tr>
<tr>
<td></td>
<td>Pohištvo</td>
<td>Najden v avtomobilu</td>
<td>Found in car dealership</td>
</tr>
</tbody>
</table>

- English COMET overall performs best
- COMET gpt-janez performs well, fails to provide alternatives
- COMET sl-gpt provides alternative
Demo: multicomet.ijs.si

• All sentences include a subject/person (PersonX)

Figure 1 Close-up of “Event-Centered” descriptor values predicted for an example Slovene sentence “PersonX is sad” (“OsebaX je žalostna” in Slovenian)

Figure 2 Close-up of “Social-Interaction” descriptor values predicted for an example Slovene sentence “John is very important” (“Janez je zelo pomemben” in Slovenian)
Discussion

• Main Contributions
  • Slovenian ATOMIC-2020 dataset, publicly available
  • Comparison of English and Slovenian commonsense models
  • Comprehensive evaluation with four metrics

• Future Work
  • Evaluate individual descriptors
  • Expand to different languages
  • Web app for 100+ languages & API

Code publicly available on GitHub