Semantic Web for Machine Translation: Challenges and Directions

Journal track

Diego Moussallem, Matthias Wauer, Axel-Cyrille Ngonga Ngomo

Department of Data Science, DICE Research Group
Paderborn University
Germany

28/10/2019
History of Machine Translation

Cold War - How was it created? First appearance

1https://medium.freecodecamp.org/a-history-of-machine-translation-from-the-cold-war-to-deep-learning
History of Machine Translation

Example of an MT idea

I
Я
ICH
YO

WANT
ХОТЕТЬ
WOLLEN
QUERER

MANY
МНОГО
VIEL
MUCHO

PERSIMMON
ХУРМА
PERSIMEONE
CAQUI

PRP, SUBJ, SINGULAR
VBP, PRESENT, SIMPLE, TRANSITIVE
JJ, DETERM, COMPARATIVE
NNS, PLURAL, COUNTABLE

---

2https://medium.freecodecamp.org/
a-history-of-machine-translation-from-the-cold-war-to-deep-learning
History of Machine Translation
Progress of MT across years

3 http://mogren.one/graphics/illustrations/2016-08-08/manning-nmt-history.png
Open Machine Translation Problems

Some challenges

- Structural divergence (order, fluency and syntactic ambiguity)
- Complex semantic ambiguity
- Pragmatic ambiguity (non standard speech - idioms)
- Data sparsity, Lack of training data
- Entities
- Out-of-vocabulary words
Background - OOV words problem

**Definition**

**OOV** : Out-of-vocabulary words are entries which do not or rarely appear in the training data and consequently the model is not able to estimate their translations.

**How and when it happens**

**Training:**
Source text EN: ...Albert Einstein died in April 16, 1955...
Reference translation PT-BR: ...Albert Einstein morreu em 16 de Abril de 1995...

**Test:**
Source text EN: ...Albert Einstein passed away in April 16, 1955...
MT output PT-BR: ...Albert Einstein **passed away** em 16 de Abril de 1995...
Background - Entity’s translation problem

Definition
When common words are entities in the target language or vice versa. Also, when the entities are very ambiguous like Kiwi which may be a fruit, a person, a computer program, and a bird depending on the language.

How and when it happens

Training:
Source text EN: ...MS Paint is a good option...
Reference Translation DE: ...Microsoft Paint ist eine gute wahl...

Test:
Source text EN: ...MS Paint is a good option...
MT output DE: ...Frau Farbe ist eine gute wahl...
Possible Solution
Knowledges Graphs in Semantic Web

4linguistic-lod.org/llod-cloud
Why Knowledge Graphs?

Figure: Knowledge Graphs

Related Work


Figure: Search Methodology

Search Engine / Digital Libraries

Google Scholar = 18,100
ACM Digital Library = 3,243
IEEE Xplore Digital Library = 57
Springer Link = 9,951
Science Direct = 2,757

Journals / Conferences / Workshops

ACL / EMTA / MT SUMMIT / AMTA / COLING / COLN / EMMNLP / ISWC / ESWC / WMT JWS / SWJ / MTJ / IJISWIS / NLE / NLLT = 242

STEP 1
Scan articles based on criteria

STEP 2
Review Abstract/Titles to include/exclude articles
Scan references to retrieve potential from articles.

STEP 3
Reviewing Full-Text
Total number of articles proposing Semantic Web in Translation Process for MT: 21

Figure: Search Methodology
## Related Work

### Machine Translation using Semantic Web - A Survey - Journal of Web Semantics

<table>
<thead>
<tr>
<th>Citation</th>
<th>Year</th>
<th>MT approach</th>
<th>SW method</th>
<th>SW resource</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. Vertan [115]</td>
<td>2004</td>
<td>EBMT</td>
<td>Annotation</td>
<td>Ontologies</td>
<td>None</td>
</tr>
<tr>
<td>N. Elita and A. Birladeanu [117]</td>
<td>2005</td>
<td>EBMT</td>
<td>SPARQL</td>
<td>Ontologies</td>
<td>None</td>
</tr>
<tr>
<td>W. Hahn and C. Vertan [116]</td>
<td>2005</td>
<td>EBMT</td>
<td>SPARQL + Annotation</td>
<td>Ontologies</td>
<td>None</td>
</tr>
<tr>
<td>C. Shi and H. Wang [118]</td>
<td>2005</td>
<td>None</td>
<td>Reasoner</td>
<td>Ontologies</td>
<td>None</td>
</tr>
<tr>
<td>N. Elita and M. Gavrilova [119]</td>
<td>2006</td>
<td>EBMT</td>
<td>SPARQL</td>
<td>Ontologies</td>
<td>None</td>
</tr>
<tr>
<td>E. Seo et al. [120]</td>
<td>2009</td>
<td>None</td>
<td>Reasoner</td>
<td>Ontologies</td>
<td>None</td>
</tr>
<tr>
<td>P. Knoth et al. [121]</td>
<td>2010</td>
<td>RBMT or SMT</td>
<td>Annotation</td>
<td>Ontologies</td>
<td>Human</td>
</tr>
<tr>
<td>A. M. Almasoud and H. S. Al-Khalifa [123]</td>
<td>2011</td>
<td>TBMT + EBMT</td>
<td>SPARQL</td>
<td>Ontologies</td>
<td>Human</td>
</tr>
<tr>
<td>L. Lesmo et al. [122]</td>
<td>2011</td>
<td>Interlingua</td>
<td>Annotation</td>
<td>Ontologies</td>
<td>Human</td>
</tr>
<tr>
<td>B. Harriehausen-Mühlbauer and T. Heuss [124]</td>
<td>2012</td>
<td>Direct</td>
<td>SPARQL + Reasoner</td>
<td>Ontologies</td>
<td>Human</td>
</tr>
<tr>
<td>K. Nebhi et al. [125]</td>
<td>2013</td>
<td>TBMT</td>
<td>Annotation</td>
<td>LOD</td>
<td>None</td>
</tr>
<tr>
<td>J. P. McCrae and P. Cimiano [126]</td>
<td>2013</td>
<td>SMT</td>
<td>Annotation</td>
<td>LOD</td>
<td>Human</td>
</tr>
<tr>
<td>D. Moussallem and R. Choren [127]</td>
<td>2015</td>
<td>SMT</td>
<td>SPARQL</td>
<td>Ontologies</td>
<td>Human</td>
</tr>
<tr>
<td>O. Lozynska and M. Davydov [128]</td>
<td>2015</td>
<td>RBMT</td>
<td>Annotation</td>
<td>Ontologies</td>
<td>Human</td>
</tr>
<tr>
<td>K. Simov et al. [129]</td>
<td>2016</td>
<td>RBMT + SMT</td>
<td>SPARQL</td>
<td>LOD</td>
<td>Automatic</td>
</tr>
<tr>
<td>T.S. Santosh Kumar. [130]</td>
<td>2016</td>
<td>SMT</td>
<td>SPARQL</td>
<td>Ontologies</td>
<td>Human</td>
</tr>
<tr>
<td>N. Abdulaziz et al. [131]</td>
<td>2016</td>
<td>SMT</td>
<td>SPARQL</td>
<td>Ontologies</td>
<td>Human</td>
</tr>
<tr>
<td>J. Du et al. [132]</td>
<td>2016</td>
<td>SMT</td>
<td>SPARQL</td>
<td>LOD</td>
<td>Automatic</td>
</tr>
<tr>
<td>A. Srivastava et al. [133]</td>
<td>2016</td>
<td>SMT</td>
<td>SPARQL + Annotation</td>
<td>LOD</td>
<td>Automatic</td>
</tr>
<tr>
<td>C. Shi et al. [134]</td>
<td>2016</td>
<td>NMT</td>
<td>Annotation</td>
<td>LOD</td>
<td>Automatic + Human</td>
</tr>
<tr>
<td>A. Srivastava et al. [135]</td>
<td>2017</td>
<td>SMT</td>
<td>SPARQL + Annotation</td>
<td>LOD</td>
<td>Automatic</td>
</tr>
</tbody>
</table>

**Figure:** Selected works
Direct Training

Related Work
Using BabelNet to Improve OOV Coverage in SMT by Du et al. 2016
Related Work

Using BabelNet to Improve OOV Coverage in SMT by Du et al. 2016

Domain Adaptation

![Diagram showing the process of using BabelNet for domain adaptation in machine translation](image)

- **Source text** is fed into BabelNet, which then maps it to a **BN:Source text**. This mapping is then used for **BN:Reference Translation**.
- The **bilingual dictionary** is used to further improve the translations.
- The ultimate output is a **Moses** machine translation system.
Post-processing

- **Source text**
- **MOSES** (statistical machine translation system)
- **Target Text**
- **Refined Target Text**
- **BabelNet**
## Final Results

<table>
<thead>
<tr>
<th>system</th>
<th>EN–PL BLEU4(%)</th>
<th>EN–PL TER(%)</th>
<th>EN–ZH BLEU4(%)</th>
<th>EN–ZH TER(%)</th>
<th>ZH–EN BLEU4(%)</th>
<th>ZH–EN TER(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>24.57</td>
<td>58.47</td>
<td>11.62</td>
<td>72.03</td>
<td>27.08</td>
<td>67.90</td>
</tr>
<tr>
<td>DTD-CLEAN-1</td>
<td>24.15</td>
<td>58.91</td>
<td>12.53</td>
<td>71.25</td>
<td>27.53</td>
<td>66.20</td>
</tr>
<tr>
<td>Best-DoAdpt API</td>
<td><strong>24.72</strong></td>
<td><strong>58.29</strong></td>
<td><strong>12.76</strong></td>
<td><strong>70.81</strong></td>
<td><strong>28.47</strong></td>
<td><strong>65.78</strong></td>
</tr>
<tr>
<td>API</td>
<td>24.71</td>
<td>58.36</td>
<td>11.71</td>
<td>71.84</td>
<td>27.23</td>
<td>67.77</td>
</tr>
</tbody>
</table>

Table 9: Comparison between BabelNet API method and others
Main Drawback
Knowledge Graphs were never learned jointly with bilingual text in the MT training phase
Outline

1. Introduction
2. MT Challenges
3. Related Work
4. Directions
   - Ontologies - Disambiguation and Non standard speech
   - Entities + Knowledge Graph Embeddings
   - Translating Knowledge Bases
5. Conclusion
Translations of Idioms relying on Knowledge Graphs

Figure: LIdioms Knowledge Graph
Translations of Idioms relying on Knowledge Graphs

Figure: Indirect Translation
Outline

1. Introduction
2. MT Challenges
3. Related Work
4. Directions
   - Ontologies - Disambiguation and Non standard speech
   - Entities + Knowledge Graph Embeddings
   - Translating Knowledge Bases
5. Conclusion
Augmenting Neural Machine Translation with Knowledge Graphs

Figure: General architecture
Directions
Entities + Knowledge Graph Embeddings

Results
Augmenting Neural Machine Translation with Knowledge Graphs

Table 1: Results in Bleu (BLEU), Met (METEOR), chrF3 on WMT newstest datasets.

<table>
<thead>
<tr>
<th>Models</th>
<th>newstest2015</th>
<th>newstest2016</th>
<th>newstest2017</th>
<th>newstest2018</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bleu  Met chrF3</td>
<td>Bleu  Met chrF3</td>
<td>Bleu  Met chrF3</td>
<td>Bleu  Met chrF3</td>
</tr>
<tr>
<td>Word-based models</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>biRNN-Lstm baseline</td>
<td>16.77 35.20 41.11</td>
<td>18.55 36.62 42.54</td>
<td>15.10 33.75 39.52</td>
<td>20.53 39.02 43.92</td>
</tr>
<tr>
<td>KG-NMT(EL+KGE)</td>
<td>19.86 38.25 42.92</td>
<td>22.38 40.40 45.18</td>
<td>18.04 36.94 41.55</td>
<td>24.87 43.49 46.88</td>
</tr>
<tr>
<td>KG-NMT(SemKGE)</td>
<td>21.49 40.19 44.72</td>
<td>24.01 42.47 46.84</td>
<td>19.66 38.89 43.11</td>
<td>27.02 45.77 48.70</td>
</tr>
<tr>
<td>CopyM models</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>biRNN-Lstm baseline</td>
<td>19.63 39.20 46.38</td>
<td>21.37 40.90 47.85</td>
<td>17.88 37.89 44.85</td>
<td>24.22 43.96 50.15</td>
</tr>
<tr>
<td>KG-NMT(EL+KGE)</td>
<td>22.46 41.67 48.28</td>
<td>25.05 44.23 50.66</td>
<td>20.77 40.58 47.04</td>
<td>28.44 47.86 53.25</td>
</tr>
<tr>
<td>KG-NMT(SemKGE)</td>
<td>24.08 43.43 49.72</td>
<td>26.70 46.08 52.05</td>
<td>22.30 42.37 48.36</td>
<td>30.55 49.92 54.71</td>
</tr>
<tr>
<td>BPE models</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>biRNN-Lstm baseline</td>
<td>15.89 36.51 45.97</td>
<td>21.95 42.88 52.68</td>
<td>16.80 39.12 49.35</td>
<td>23.85 45.85 54.98</td>
</tr>
<tr>
<td>KG-NMT(EL+KGE)</td>
<td>N/A N/A N/A</td>
<td>N/A N/A N/A</td>
<td>N/A N/A N/A</td>
<td>N/A N/A N/A</td>
</tr>
<tr>
<td>KG-NMT(SemKGE)</td>
<td>21.74 41.41 50.04</td>
<td>24.86 44.32 53.59</td>
<td>20.45 40.62 49.45</td>
<td>28.02 47.51 55.16</td>
</tr>
</tbody>
</table>
Outline

1 Introduction

2 MT Challenges

3 Related Work

4 Directions
   • Ontologies - Disambiguation and Non standard speech
   • Entities + Knowledge Graph Embeddings
   • Translating Knowledge Bases

5 Conclusion
Enrichment of Knowledge Bases

**Training**
- **1.** Source KB
- **2.** Target KB
  - Extract Bilingual Content
  - Bilingual Knowledge
  - Extract Aligned Triples

**Translation**
- **2.** Bilingual Generic Corpora
  - Source Text
  - Target Text
  - Insert External Knowledge
- **3.** Text-based NN model
- **3.** Triple-based NN model

- **2.** Enriched Target KB

**Figure:** THOTH architecture

*Diego Moussallem, Matthias Wauer, Axel-CySemantic Web for Machine Translation: Cha*
Directions
Translating Knowledge Bases

Results
Enrichment of Knowledge Bases

![Accuracy Chart]

**Fig. 2: Accuracy overall**
Summary and Future Work

Summary
- Some MT challenges
- How to use Semantic Web in MT

What Next?
- Access more than one KGs at once
- Use ontologies as MT features
- Investigate different Neural Networks with KGs
Thank you for your Attention!

Diego Moussallem
diego.moussallem@upb.de
https://go.upb.de/moussallem
Follow us on Twitter @DiceResearch
Follow me on Twitter @diegomoussallem

⇒ GitHub: https://github.com/dice-group
⇒ Website:
    https://dice.cs.uni-paderborn.de