Cross-lingual Knowledge Projection Using Machine Translation and Target-side Knowledge Base Completion

Naoki Otani[1] Hirokazu Kiyomaru[2] Daisuke Kawahara[2] Sadao Kurohashi[2]
[1]Carnegie Mellon University, [2]Kyoto University

Translated 18,747 facts (tuples) of commonsense knowledge with high precision.
Addressed the problem of projection ambiguity by combining MT and KBC.

Existing Japanese facts 69,902

Background - Commonsense Knowledge

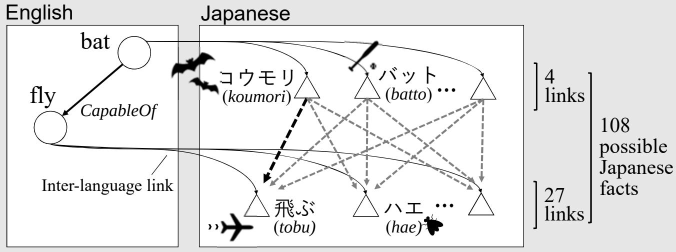
Things that every person should know.
Important to understand human languages.

ConceptNet (Speer et al., 2017)
The largest multi-lingual knowledge base of commonsense
- Tuples (facts) of commonsense (bat, *CapableOf*, fly)
- Nodes are represented in undisambiguated words/phrases

Problem – Large gap between English and other languages
Unique English facts: 2,828,394
Unique Japanese facts: 69,902 (~2:5%)

Task

Projecting English facts into other languages.
Challenge: projection of commonsense is ambiguous.



Problem Setting

f^s: fact in English, $f_1^t, \cdots, f_n^t$: projection candidates in a target language

Goal: find the most appropriate fact by $\hat{f}^t = \mathrm{argmax}_i\ h(f_i^t | f^s)$

Our Approach – Combining Machine Translation and Target-side Knowledge Base Completion

Machine Translation (MT)

Calculating trans. probs. with an off-the-shelf neural MT model
Implementation: lamtram (Graham, 2015) + BPE (Sennrich et al., 2016)

$$x_{MT}\big((\text{koumori}, \textit{CapableOf}, \text{tobu}) \mid (\text{bat}, \textit{CapableOf}, \text{fly})\big)$$

★ $\boldsymbol{e_1}$ wa $\boldsymbol{e_2}$ koto ga dekiru . $\quad$ $\boldsymbol{e_1}$ can $\boldsymbol{e_2}$.

$$= \big(P(\textbf{koumori}\text{ wa }\textbf{tobu}\text{ koto ga dekiru . } | \text{ A }\textbf{bat}\text{ can }\textbf{fly}\text{ .})\big)^{\frac{1}{7}}$$

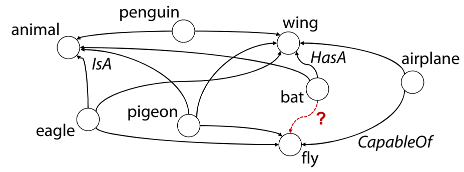
$$= \begin{pmatrix} P(\text{koumori}|\text{A bat}\cdots.) \times P(\text{wa}|\text{koumori, A bat}\cdots.) \\ \times \cdots \times P(.|\text{dekiru}, \cdots, \text{koumori, A bat}\cdots.) \end{pmatrix}^{\frac{1}{7}}$$

★ *Converting facts into sentences based on rules*

Relation	e_1, e_2	English	Japanese	Chinese
AtLocation	NP, NP	You are likely to find e_1 in e_2 .	e_2 de e_1 wo miru koto ga dekiru .	Ni keyi zai e_2 zhaodao e_1 .
CapableOf	NP, VP	e_1 can e_2 .	e_1 wa e_2 koto ga dekiru .	e_1 hui e_2
MadeOf	NP, NP	e_1 is made of e_2 .	e_1 wa e_2 kara tsukurareru .	e_1 ke yi yong e_2 zhi cheng .

Knowledge Base Completion (KBC)

Evaluate the plausibility of a target-side fact based on existing information in a knowledge base.



Bilinear model (Li et al., 2017)

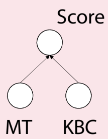
$$x_{KBC}((\text{koumori}, \textit{CapableOf}, \text{tobu})) = \sigma(\boldsymbol{u}_{\text{koumori}}^T W_{\textit{CapableOf}}\, \boldsymbol{u}_{\text{tobu}})$$

Node vector: $\boldsymbol{u} = \tanh(W\boldsymbol{v} + \boldsymbol{b}) \in \mathbb{R}^d$,
$\boldsymbol{v} \in \mathbb{R}^{d'}$: word vector, $W, \boldsymbol{b}$: parameters
Relation matrix: $W^{d \times d}$
The model parameters $(W, \boldsymbol{b})$ are learned to minimize a cross-entropy loss on training facts.

Combination – Two Simple Methods

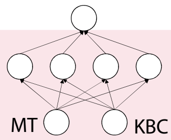
1. Linear transformation (LIN)

$h(x) = w_r^T x + b_r, w_r \in \mathbb{R}^2, b_r \in \mathbb{R}$
$(x = (x_{MT}, x_{KBC})$: scores, r: relation)



2. Multi-layer Perceptron (MLP)

$h(x) = w_r^{(2)^T} z(x) + b_r^{(2)}$
$z(x) = \tanh(W^{(1)^T} x + b^{(1)})$
$W^{(1)} \in \mathbb{R}^{2 \times c}, b^{(1)} \in \mathbb{R}^c, w^{(2)} \in \mathbb{R}^c, b_r^{(2)} \in \mathbb{R}$



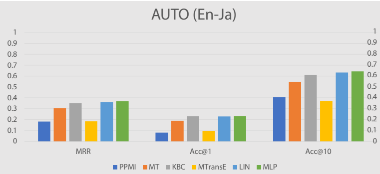
Experiments

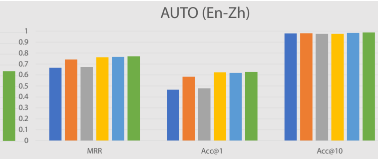
Data source: ConceptNet 5.5.0 (Speer et al., 2017)
Two evaluation sets:
- *AUTO:* large, automatically collected fact alignments
- *MANUAL*: small, manually verified fact alignments

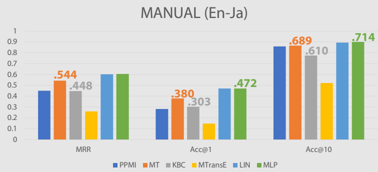
Evaluation metrics:
- Mean reciprocal rank (MRR), top-k accuracy (Acc@k)

Baselines:
- PPMI / MT / KBC / MTransE (Chen et al., 2017)

Our methods: LIN / MLP





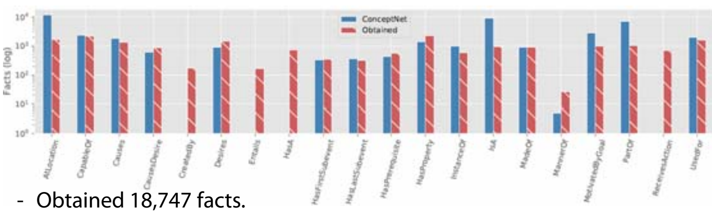


(**blue**, *RelatedTo*, color)
MT, MLP
MT was more accurate at *RelatedTo* relation.
[blue] color
(**ao**, *RelatedTo*, shikisai)

(**rock**, *IsA*, music)
MT KBC, MLP
[stone] music [music] music
(**iwa**, *IsA*, myusikku) (**rokku**, *IsA*, myusikku)
Pre-trained word emb. of "rokku" was similar to those of other music genres (e.g. jazz.)

Rapidly Acquiring Japanese Commonsense with the Proposed Method + Crowdsourcing

1. We projected 10k English facts covering 20 relation types into Japanese
2. To further improve the quality, we verified the top-10 predictions of MLP using crowdsourcing
   - *Screening top-10 is fast.* – 838 workers and 25 hours



- Obtained 18,747 facts.
- Equivalent or larger amount of knowledge for 12 relation types.

- Chen et al. 2017. Multi-lingual Knowledge Graph Embeddings for Cross-lingual Knowledge Alignment. In *IJCAI*.
- Graham. 2015. lamtram: A Toolkit for Language and Translation Modeling using Neural Networks.
- Li et al. 2016. Commonsense Knowledge Base Completion. In *ACL*.
- Sennrich et al. 2016. Improving Neural Machine Translation Models with Monolingual Data. In *ACL*.
- Speer et al. 2017. ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. In *AAAI*.

Contact: Naoki Otani <notani@cs.cmu.edu>
Code&Data: https://github.com/notani/CLKP-MTKBC