Multi-target Machine Translation with Multi-synchronous Context Free Grammars

Graham Neubig, Philip Arthur, Kevin Duh
Nara Institute of Science and Technology (NAIST), Japan

Motivation

- When translating into language T1, equivalent translations into a second language T2 can help

Source	??	Target 1 (T1)	Target 2 (T2)
اتفاقية كيوتو		京都条约	Kyoto Treaty
		京都制裁	Kyoto Sanction
		京都议定书	Kyoto Protocol

- In MT, if *T1 has a weak language model*, can we use a *strong language model in T2* to improve results?

Proposed Framework

- Build on the well-known *synchronous context-free grammars (SCFG)*

I' ONU → 联合国
Kyoto → 京都
X_0 a ratifié X_1 → X_0 批准 了 X_1
le Protocole de X_0 → X_0 议定书

- Propose *multi-synchronous context-free grammars (MSCFGs)*, with multiple targets

I' ONU → 联合国 | the UN
Kyoto → 京都 | Kyoto
X_0 a ratifié X_1
→ X_0 批准 了 X_1 | X_0 ratified the X_1
le Protocole de X_0
→ X_0 议定书 | the X_0 Protocol

Learning MSCFGs

- Learn from *tri-lingual parallel data*

1) Alignment

联合国 批准 了 京都 议定书
I' ONU a ratifié le Protocole de Kyoto
the UN ratified the Kyoto Protocol
Independent

2) Phrase Extraction

ONU → 联合国
a → 了
ratifié → 批准
a ratifié → 批准 了
…

(1) ONU → UN
(2) a ratifié → ratified
…

(1) ONU → 联合国 | UN
(2) a ratifié → 批准 了 | ratified
…

3) Calculate Features

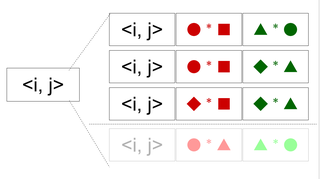
Standard:
Translation Prob. $\sum_{\langle f,e_1\rangle} \log P(f|e_1)$ $\sum_{\langle f,e_1\rangle} \log P(e_1|f)$
Lexical Prob. $\sum_{\langle f,e_1\rangle} \log P_{lex}(f|e_1)$ $\sum_{\langle f,e_1\rangle} \log P_{lex}(e_1|f)$
Words $|E_1|$
LM $\log P(E_1)$
Rules $|D|$

Additional:
Translation Prob. $\sum_{\langle f,e_2\rangle} \log P(f|e_2)$ $\sum_{\langle f,e_{1,}e_2\rangle} \log P(f|e_1,e_2)$ $\sum_{\langle f,e_2\rangle} \log P(e_2|f)$ $\sum_{\langle f,e_{1,}e_2\rangle} \log P(e_1,e_2|f)$
Lexical Prob. $\sum_{\langle f,e_2\rangle} \log P_{lex}(f|e_2)$ $\sum_{\langle f,e_2\rangle} \log P_{lex}(e_2|f)$
Words $|E_2|$
LM $\log P(E_2)$

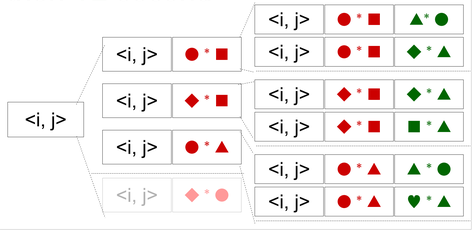
Decoding w/ MSCFGs

- Two ways to handle increased search space due to two language models

Joint decoding: T1 and T2 simultaneously



Sequential decoding: T1 first then T2 second



Experiments

MultiUN Corpus:
Parallel, T1 LM data: 100,000 Sentences
T2 LM data: 4,000,000 Sentences
S: en T1, T2: ar, es, fr, ru, zh (all combinations)

Baseline: SCFG-based 1-target Hiero Grammar

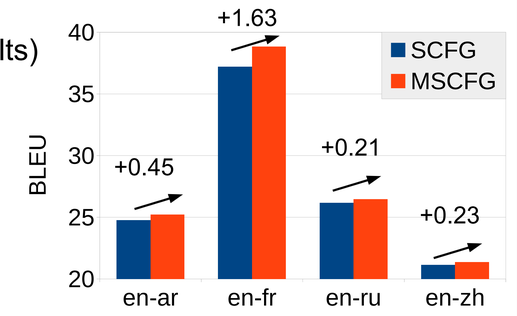
Proposed: MSCFG-based 2-target Hiero Grammar

Result 1: Does second target Help?

e.g. T2=es (best results)

Answer:
Yes! In most cases accuracy improves

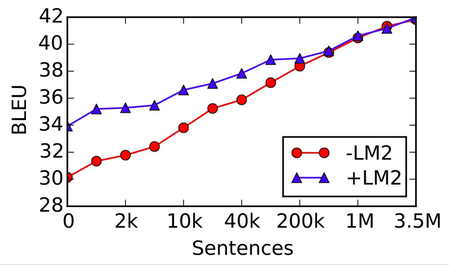
Particularly effective in similar languages



Result 2: Influence of T1 LM strength?

e.g. T1=fr, T2=es

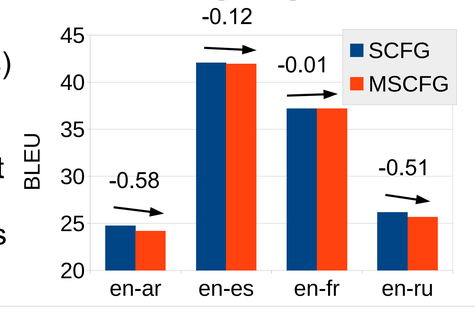
Answer:
As expected, works best when T1 LM is less strong:
Covers weakness of T1



Result 3: Influence of T2 language?

e.g. T2=zh (worst results)

Answer:
When T2 is very different results less good, due to rule extraction constrants



Results on decoding, model size, etc. in paper

Try out the code/scripts!
http://phontron.com/project/naacl2015