Integrating Weakly Supervised Word Sense Disambiguation Into Neural Machine Translation

https://github.com/idiap/sense_aware_NMT

Xiao Pu [1], Nikolaos Pappas [2], James Henderson [2] and Andrei Popescu-Belis [3]

[1] Nuance Communications, Germany [2] Idiap Research Institute, Switzerland [3] HEIG-VD / HES-SO, Switzerland

Problem

The correct translation of polysemous words remains a challenge for machine translation.

Example

Source sentence: *How can we guarantee the system of prior notification for high-risk products at ports that have the necessary facilities to **deal** with them?*

Reference translation: *Comment pouvons-nous garantir le système de notification préalable pour les produits présentant un risque élevé dans les ports qui disposent des installations nécessaires pour **traiter** ces produits ?*

Baseline neural MT: *[…] qui disposent des moyens nécessaires pour y **remédier** ?*

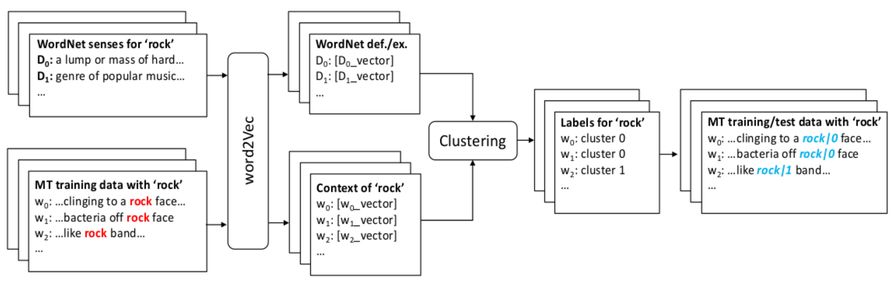
Sense-aware neural MT: *[…] qui disposent des installations nécessaires pour les **traiter** ?*

Solution

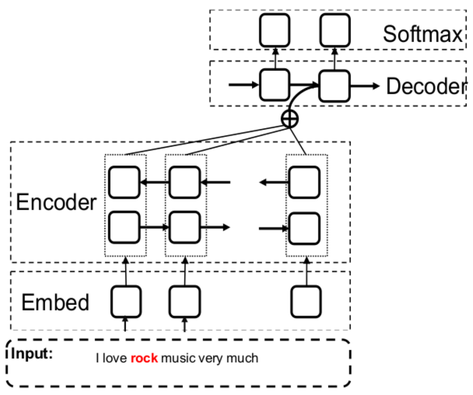
We demonstrate that the explicit modeling of word senses can help NMT by using *combined vector representations of word types and senses*, which are inferred from contexts that are larger than those used in state-of-the-art NMT systems.

Architecture of WSD and NMT

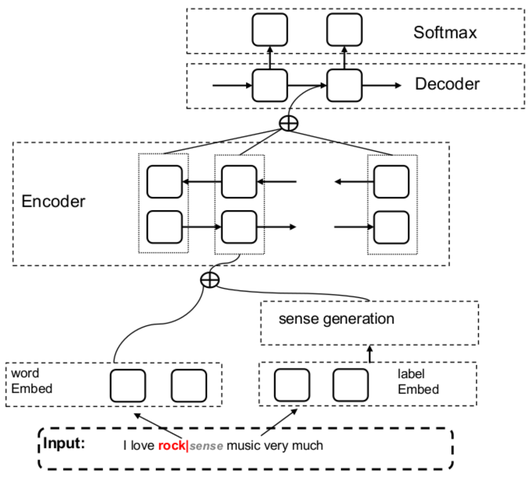
Adaptive word sense disambiguation



RNN seq2seq neural MT



Sense-aware neural MT

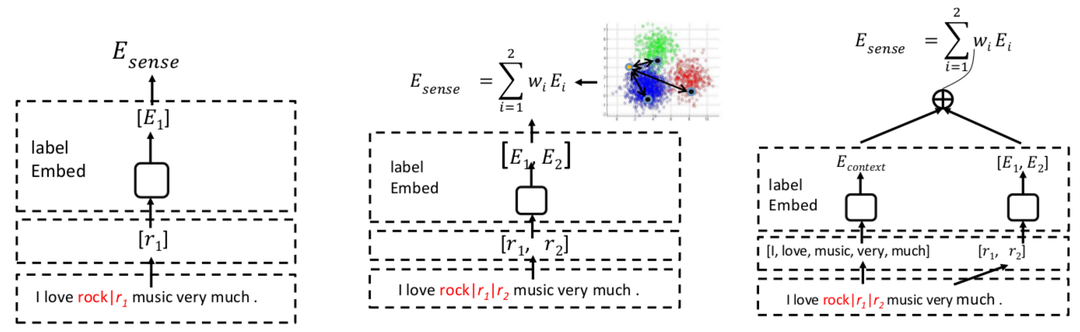


Datasets: from English to five target languages

TL	Train sent.	Dev. sent.	Test sent.	Labels		Words
				Noun	Verbs	
FR	0.5M	5k	50k	3910	1627	2006
	5.3M	4.6k	6k	8276	3059	3876
DE	0.5M	5k	50k	3885	1576	1976
	4.5M	3k	5k	7520	1634	3194
ES	0.5M	5k	50k	3862	1627	1987
	3.9M	4.6k	6k	7549	2798	3558
ZH	0.5M	5k	50k	3844	1475	1915
NL	0.5M	5k	50k	3915	1647	2210

Sense-aware Neural MT

1. Top sense (TOP): directly use the word sense selected by WSD.
2. Weighted average of sense (AVG): consider all listed senses and the respective cluster centroids learned by WSD.
3. Attention-based sense weight (ATT): compute the relatedness probability for each listed senses dynamically during encoding.
4. Attention-based sense weight with initialization of embeddings (ATT_{ini}): similar to ATT model but initialize the embeddings of source word dictionary using word2vec vectors.



Results

	EN/FR	EN/DE	EN/ZH	EN/ES	EN/NL
SMT baseline	31.96	20.78	23.25	39.95	23.56
Graph	32.01 (+ .05)	21.17 (+ .39)	23.47 (+ .22)	40.15 (+ .20)	23.74 (+ .18)
CRP	32.08 (+ .12)	21.19 (+ .41)	23.55 (+ .29)	40.14 (+ .19)	23.79 (+ .23)
k-means	32.20 (+ .24)	21.32 (+ .54)	23.69 (+ .44)	40.37 (+ .42)	23.84 (+ .26)
NMT baseline	34.60	25.80	27.07	44.09	24.79
K-means + TOP	24.52 (- .08)	25.84 (+ .04)	26.93 (- .14)	44.14 (+ .05)	24.71 (- .08)
K-means + AVG	35.17 (+ .57)	26.47 (+ .67)	27.44 (+ .37)	45.05 (+ .97)	25.04 (+ .25)
None + ATT	35.32 (+ .72)	26.50 (+ .70)	27.56 (+ .49)	44.93 (+ .84)	25.36 (+ .57)
K-means + ATT_{ini}	35.78 (+ 1.18)	26.74 (+ .94)	27.84 (+ .77)	45.18 (+ 1.09)	25.65 (+ .86)

BLEU scores of sense-aware NMT systems: ATT_{ini} is the best one among SMT and NMT systems.

	EN/FR		EN/ES	
	NT12	NT13	NT12	NT13
Baseline	29,09	29,60	32,66	29,57
None + ATT	29,47 (+.38)	30,21 (+.61)	33.15 (+.49)	30,27 (+.70)
K-means + ATT_{ini}	**30.26** (+1.17)	**30.95** (+ 1.35)	**34.14** (+ 1.48)	**30.67** (+ 1.1)

BLEU scores on WMT NewsTest (NT) 2012 and 2013 test sets for two language pairs

NMT model	NT14	NT15
Context-dependent (Choi et al., 2017)	-	21.99
Context-aware (Zhang et al., 2017)	22.57	-
Self-attentive (Werlen et al., 2018)	23.2	25.5
Baseline	22.79	24.94
None + ATT	23.34	25.28
K-means + ATT_{ini}	**23.85 (+1.14)**	**25.71 (+0.77)**

BLEU scores for **EN/DE** translation on WMT NewsTest (NT) 2014 and 2015

Conclusion

1. Proposed a neural MT system enhanced with an attention-based model for representing multiple word senses, making use of a larger context to disambiguate words that have multiple possible translations.
2. Proposed several adaptive context-dependent clustering algorithms for WSD and several ways to combine them with NMT.

Acknowledgements

This work was supported by the Swiss National Science Foundation (SNSF) through the Sinergia project **MODERN: Modeling Discourse Entities and Relations for Coherent Machine Translation** (grant n. 147653, www.idiap.ch/project/modern) and by the European Union through the Horizon 2020 project **SUMMA: Scalable Understanding of Multilingual Media** (grant n. 688139, www.summa-project.eu).

References

- Xiao Pu, Nikolaos Pappas, and Andrei Popescu-Belis. 2017. Sense-aware statistical machine translation using adaptive context-dependent clustering. In *Proc. of the 2nd Conference on Machine Translation*, Copenhagen, p. 1-10.
- Eneko Agirre and Philip Edmonds. 2007. Word Sense Disambiguation: Algorithms and Applications. *Springer-Verlag*, Berlin, Germany.
- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *Proc. of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL),* pages 33-41, Athens, Greece.