Generating and Understanding Color References with Bilingual Multi-Task Learning

Anonymous ACL submission

Abstract

We show that training on English data improves modeling of context-sensitive color description generation in Chinese on a reference game task. Furthermore, a Chinese description understanding model derived by Bayesian inference from this bilingually-trained speaker outperforms simple classifiers, both monolingual and bilingual.

1 Introduction

Lexical semantic spaces differ among languages; English and Chinese have different color terms. Do speakers of different languages use context in different ways?

(Why colors? They are a well-understood referent space, allow for illustrative analysis.)

Training on English data improves Chinese referring expression generation. This suggests the ability to transfer either context sensitivity or compositionality between the two languages.

2 Task

Reference game.

Speaker and listener sides of the task.

3 Data collection

Hawkins (2015)

4 Human data analysis

Length, specificity.

Comparatives, superlatives, negation.

Success rates.

5 Models

Monolingual L_0 (separately trained for en, zh): Gaussian listener in Monroe et al. (2017).

Logistic regression baseline.

Test data	en		zh	
Train data	en	en+zh	zh	en+zh
S (ppl)	20.37	28.65	67.75	64.24
L (acc %)	83.30	82.64	64.98	64.86
$L(S)$	80.51	79.01	64.20	67.27

Table 1: Speaker perplexities (top) and listener accuracies (bottom).

Bilingual L_0 (trained on both en and zh data). Three designs: no change from Gaussian listener, addition of linear transform on word vectors, use of pretrained Glove and Zou word vectors.

Monolingual S_0 (separately trained): context-to-sequence RNN speaker from Monroe et al. (2017) based on color-to-sequence speaker from Monroe et al. (2016).

Bilingual S_0 (trained on both en and zh data), with or without an extra bit specifying whether the output should be in English or Chinese.

6 Experimental results

Bilingual S0 is better for Chinese that a monolingual one (with the bit to say which language to use).

Bilingual L(S) is better for Chinese than a bilingual L0.

Bilingual L(S) is better for Chinese than all monolingual listeners.

7 Model analysis

Split model results by condition, to measure effect of needing the context. Can we identify sentences that are syntactically similar, to see how much of the transfer is syntactic as opposed to pragmatic?

8 Related work

Collobert and Weston (2008)

一	个	是	浅	紫色			
yi	ge	shì	qiǎn	zǐsè			
one	CL	is	shallow	purple			
一	个	是	艳	紫色			
yi	ge	shì	yàn	zǐsè			
one	CL	is	bright	purple			
剩下	的	那个	色	就是	要	选	的
shèngxia	de	nàge	sè	jiùshì	yào	xuǎn	de
remain	DE	that	color	is	want	choose	DE

"One is pale purple, one is bright purple. The remaining color is the one to choose."

Figure 1: This is just to make sure LaTeX will display Chinese text correctly.

Johnson et al. (2016)
Wu et al. (2016)
Kaiser et al. (2017)

9 Conclusion

References

Ronan Collobert and Jason Weston. 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th International Conference on Machine Learning*. ACM, pages 160–167. https://wiki.inf.ed.ac.uk/twiki/pub/CSTR/Speak11To12Semester2/collobert-2008.pdf.

Robert X. D. Hawkins. 2015. Conducting real-time multiplayer experiments on the web. *Behavior Research Methods* 47(4):966–976.

Melvin Johnson, Mike Schuster, Quoc V Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, et al. 2016. Google's multilingual neural machine translation system: enabling zero-shot translation. *arXiv preprint arXiv:1611.04558* https://arxiv.org/abs/1611.04558.

Łukasz Kaiser, Aidan N Gomez, Noam Shazeer, Ashish Vaswani, Niki Parmar, Llion Jones, and Jakob Uszkoreit. 2017. One model to learn them all. *arXiv preprint arXiv:1706.05137* https://arxiv.org/abs/1706.05137.

Will Monroe, Noah D. Goodman, and Christopher Potts. 2016. Learning to generate compositional color descriptions. In *Proceedings of the 2016 Conference on Empirical Methods on Natural Language Processing (EMNLP)*. pages 2243–2248.

Will Monroe, Robert X.D. Hawkins, Noah D. Goodman, and Christopher Potts. 2017. Colors in context: A pragmatic neural model for grounded language understanding. *TACL* 5:325–338. https://nlp.stanford.edu/pubs/monroe2017colors.pdf.

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, et al. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144* https://arxiv.org/pdf/1609.08144.pdf.

A Model details

Tuning method and final hyperparameters
Vocab sizes