Appendix: Bridging the Defined and the Defining: Exploiting Implicit Lexical Semantic Relations in Definition Modeling

Koki Washio[*†1] **Satoshi Sekine**[†2] **Tsuneaki Kato**[*3]
[*]Department of Language and Information Sciences, The University of Tokyo
[†]RIKEN Center for Advanced Intelligence Project
{[1]kokiwashio@g.ecc, [3]kato@boz.c}.u-tokyo.ac.jp
[2]satoshi.sekine@riken.jp

Name	Value
Word Embedding	
- Google Word2Vec	300 dimensions
Definition Encoder	
- 1-layer LSTM	300 dimensions
Mini-batch size	32
Optimizer	Adam
Learning rate	3e-4
Epoch	50
Gradient clip threshold	5.0
Consistency penalty weight	{1, 2, 4, 8, 16, 32, 64}

Table 1: Hyperparameters for definition embeddings

Name	Value
Word Embedding	
- Google Word2Vec	300 dimensions
Character Embedding	20 dimensions
Character CNN	20 dimensions
Definition Decoder	
- 2-layer LSTM	300 dimensions
Mini-batch size	64
Dropout rate	0.5
Optimizer	Adam
Learning rate	1.2e-6
Gradient clip threshold	5.0

Table 2: Hyperparameters for context-agnostic definition generation

1 Hyperparameters for Definition Embeddings

Table 1 shows the hyperparameters in our experiments for definition embeddings. In our experiments for definition embeddings, we implemented CPAE with Pytorch[1]. The settings used in this study are the same as Bosc and Vincent (2018), except for the initialization of the word embeddings with Google Word2Vec.

We used the dataset[2] extracted from WordNet (Fellbaum, 1998) for the training. We trained the model for 50 epochs. The model was saved every five epochs. The consistency penalty weight for CPAE varies among the following values, $\{1, 2, 4, 8, 16, 32, 64\}$, and the value was chosen in the model selection phase.

We used the development set of SimVerb(Gerz et al., 2016) and MEN(Bruni et al., 2014) for the hyperparameter tuning and model selection. Based on the previous work (Bosc and Vincent, 2018), we select the best model according to the weighted mean score that weights SimVerb twice as MEN.

[1]https://pytorch.org/

[2]https://github.com/tombosc/cpae

Name	Value
Word Embedding	
- Google Word2Vec	300 dimensions
Attention MLP	300 dimentions
Definition Decoder	
- 2-layer LSTM	300 dimensions
Mini-batch size	64
Dropout rate	0.5
Optimizer	Adam
Learning rate	1.2e-6
Gradient clip threshold	5.0

Table 3: Hyperparameters for context-aware definition generation

2 Hyperparameters for Definition Generation

Table 2 and Table 3 show the hyperparameters for our experiments for the context-agnostic and context-aware definition generation, respectively. We implemented S+G+CH+HE and S+I-Attention with Pytorch. We did not update the word embeddings during the training. This improved the overall performance.

The training was stopped if the perplexity on the development set did not improve for five epochs. The best model was chosen based on the perplexity on the development set.

References

Tom Bosc and Pascal Vincent. 2018. Auto-encoding dictionary definitions into consistent word embeddings. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1522–1532, Brussels, Belgium. Association for Computational Linguistics.

Elia Bruni, Nam-Khanh Tran, and Marco Baroni. 2014. Multimodal distributional semantics. *Journal of Artificial Intelligence Research*, 49:1–47.

Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, Mass.

Daniela Gerz, Ivan Vulić, Felix Hill, Roi Reichart, and Anna Korhonen. 2016. SimVerb-3500: A large-scale evaluation set of verb similarity. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2173–2182, Austin, Texas. Association for Computational Linguistics.