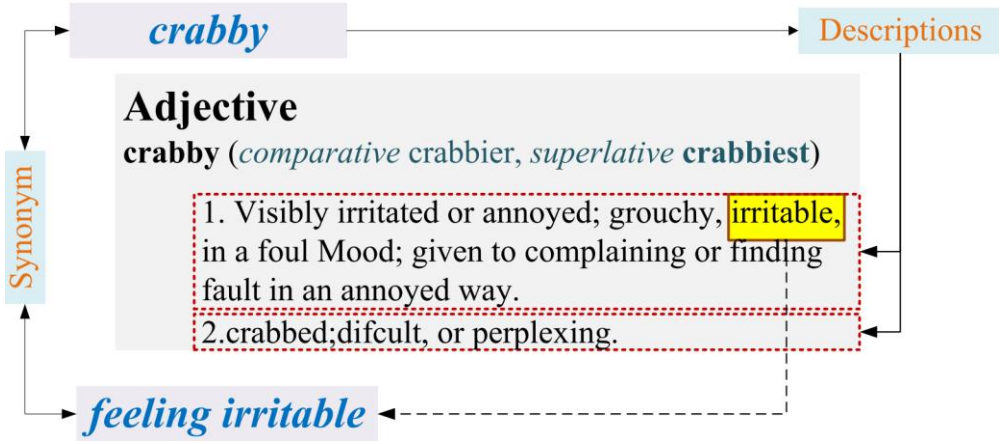
ProSyno: Context-Free Prompt Learning for Synonym Discovery

Song Zhang, Lei He, Dong Wang, Hongyun Bao, Suncong Zheng, Yuqiao, Liu, Baihua Xiao, Jiayue Li, Dongyuan Lu, Nan Zheng

Frontiers of Computer Science, DOI: 10.1007/s11704-024-3900-z

Problems & Ideas

- Problems of conventional synonym discovery approaches:
  - It's easy to make false positive predictions just relying on pre-trained embeddings.
  - Knowledge graphs and context may not be available in some domains, hindering models from generalizing to these fields.
- Ideas: A context-free prompt learning based synonym discovery method which takes the world's largest freely available dictionary Wiktionary as a semantic source.



A sample which shows that a word description in Wiktionary helps to distinguish synonym.

Main Contributions

- Contributions:
  - Descriptions in the largest freely available dictionary, i.e., Wiktionary, are integrated, so that the model can mitigate the semantic gap between term pairs and get rid of the dependency on contexts and KGs;
  - To dynamically obtain a highly informative representation from multiple descriptions of a word, a dynamical matching mechanism is designed to fuse them through the matching degree with the candidate term;
  - The method is the first try to introduce the idea of prompting into the context-free synonym discovery task, which enable our model to generalize to other datasets without any annotated data.

Datasets	AAP	TwADR-L	CADEC	ANV
WordCNN	81.41	44.78	-	-
WordGRU	85.71	-	-	-
BERT	87.46	47.02	-	-
BioBERT	88.39	48.32	-	-
CODER	-	-	59.01	-
MoE-ASD	87.30	47.65	58.24	92.16
ProSyno	**90.22**	**51.49**	**60.55**	**94.43**

Datasets	TwADR-L	CADEC	ANV
WordCNN	44.78	-	-
WordGRU	-	-	-
BERT	47.02	-	-
BioBERT	48.32	-	-
CODER	-	**59.01**	-
MoE-ASD	47.65	58.24	92.16
ProSyno-AA	**50.03**	58.92	**93.57**

Left: Comparisons of ProSyno against state-of-the-art performances; Right: Performance of generalization (%)