ParaAMR: A Large-Scale Syntactically Diverse Paraphrase Dataset by AMR Back-Translation

Kuan-Hao Huang[1] Varun Iyer[2] I-Hung Hsu[3] Anoop Kumar[4] Kai-Wei Chang[1,4] Aram Galstyan[3,4]

[1] University of California, Los Angeles [2] University of Illinois Chicago
[3] Information Science Institute, University of Southern California [4] Amazon Alexa AI

Paraphrase Generation

Paraphrase generation benefits many NLP applications

- Question answering
- Chatbots
- Data augmentation
- Robustness

Challenge: Large-Scale High Quality Paraphrase Data

Human-annotated dataset

- MRPC, PAN, Quora
- High quality but limited scale

Automatically generated dataset

- ParaNMT, ParaBank 1, ParaBank 2 (back-translation)
- Large scale but lack of syntactic diversity

ParaAMR

https://github.com/uclanlp/ParaAMR

Generate Syntactically Diverse Paraphrases by AMR Back-Translation

Abstract Meaning Representations (AMR)

- A directed graph capturing the abstract meaning of a sentence
- Nodes represent semantic concepts
- Edges represent semantic relations
- Focus (root node) represents main assertion of the sentence

Quantitative Analysis

Automatic Scores

Dataset	Semantic Similarity (↑)	Lexical Diversity 1 - BLEU (↑)	Lexical Diversity 1 - ∩/∪ (↑)	Syntactic Diversity TED-3 (↑)	Syntactic Diversity TED-F (↑)
ParaNMT (Wieting and Gimpel, 2018)	**84.28**	70.71	45.78	3.28	13.94
ParaBank1 (Hu et al., 2019a)	81.77	78.19	52.59	3.59	14.53
ParaBank2 (Hu et al., 2019b)	82.50	**88.82**	**59.61**	4.04	17.41
ParaAMR (Ours)	82.05	87.86	53.10	**5.86**	**22.07**

Human Evaluation Scores

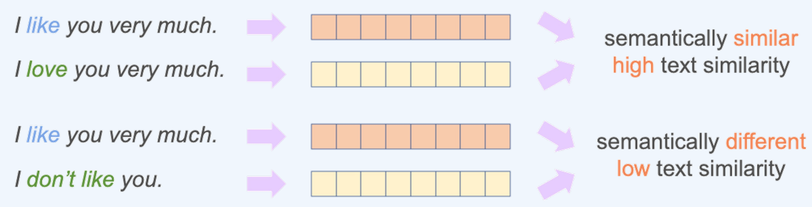
Datasets	Semantic Similarity 3(%)	2(%)	1(%)	Average	Syntactic Diversity 3(%)	2(%)	1(%)	Average
ParaNMT (Wieting and Gimpel, 2018)	28.7	46.7	24.6	**2.04**	16.7	45.0	38.3	1.78
ParaBank1 (Hu et al., 2019a)	26.8	49.0	24.2	2.03	15.1	47.8	37.1	1.78
ParaBank2 (Hu et al., 2019b)	26.8	50.3	22.9	**2.04**	14.2	51.8	34.0	1.80
ParaAMR (Ours)	26.5	47.2	26.3	2.00	18.2	53.8	28.0	**1.90**

Qualitative Examples

Source Sentence	I know for them to approve this price, they'll need statistical documentation.
ParaNMT	I know that in order to accept this award, they'll need a statistical analysis.
ParaBank1	I know that to accept this prize, they're going to need statistical analysis. I know that in order to accept this prize, they're going to need a statistic analysis. I know that if they accept this prize, they're gonna need a statistical analysis.
ParaBank2	I know that to accept that prize, they're going to need a statistical analysis. I know that in order to accept this prize, they will require a statistical analysis. I know they'll require statistical analysis to accept that prize.
ParaAMR	I know they need statistical documentation to approve this price. There is statistic documentation I know they need to approve these prices. They need statistical documentation to approve these prices, I know.

Applications

Learning Sentence Embeddings



Dataset	Pearson's r	Spearman's r
ParaNMT	74.38 ± 0.70	73.80 ± 0.42
ParaBank1	74.80 ± 1.33	74.56 ± 1.02
ParaBank2	75.39 ± 0.29	75.17 ± 0.25
ParaAMR (ours)	**77.70** ± 0.40	**75.72** ± 0.43

Syntactically Controlled Paraphrase Generation

Dataset	Quora	MRPC	PAN
ParaNMT	47.38 ± 0.39	45.24 ± 0.61	39.45 ± 0.50
ParaBank1	46.21 ± 0.26	44.52 ± 0.18	39.85 ± 0.11
ParaBank2	46.86 ± 0.45	45.17 ± 0.39	40.20 ± 0.56
ParaAMR (ours)	**48.50** ± 0.11	**47.38** ± 0.19	**40.30** ± 0.10

Data Augmentation for Few-Shot Learning

Dataset	MRPC	QQP	RTE
15-Shot Learning			
15-Shot Baseline	59.93	63.18	54.05
ParaNMT	49.26	63.54	**55.68**
ParaBank1	59.56	63.72	54.59
ParaBank2	58.46	63.54	54.05
ParaAMR (ours)	**62.87**	**64.08**	52.97
30-Shot Learning			
30-Shot Baseline	68.38	64.93	54.51
ParaNMT	67.65	66.20	52.71
ParaBank1	64.46	64.86	53.79
ParaBank2	68.38	64.91	54.15
ParaAMR (ours)	**69.36**	**67.03**	**55.60**