Contributions
- A Supervised Contrastive Learning framework for prompt-based few-shot learners
- An effective data augmentation method using prompts for contrastive learning with prompt-based learners.

Overview

Algorithm 1 Our method
1. Max Step = 1000,
2. LM: Language model,
3. Train Set, Training set,
4. Sample: Randomly sampling function,
5. Concatenate: The function to concatenate two strings.
6. CE: Cross Entropy loss
7. SupCon: Supervised Contrastive loss.
8. Prompt-based fine tuning with demonstrations to help model to know what is “great” / “terrible”.
9. learning from MML Loss
10. output = LM(input)
11. L_MML = CE(output, y)
12. L_MML backward()
13. optimizer.step()
14. prompts from SupCon loss
15. output = SupCon(output, input)
16. L_SupCon backward()
17. optimizer.step()
18. end for

Overview of our proposed method.

Contrastive Learning for Prompt-based Few-shot Language Learners

Yiren Jian, Chongyang Gao, Soroush Vosoughi

Algorithm

With demonstrations vs. Without demonstrations

Different kinds of fine tuning

Conventional fine tuning, classifying at hidden states of [CLS].
- Prompt-based fine tuning, predicting the best suited missing word.
- Prompt-based fine tuning with demonstrations to help model to know what is “great” / “terrible”.

Random templates/demonstrations vs.

Comparing our random templates/demonstrations as data augmentation to Synonym Replacement (SR), Random Insertion (RI), Random Swapping (RS), Random Deletion (RD) and EDA [4].

Overview of our proposed method.

The average improvements achieved by our method on the top K hardest tasks.

Algorithm

With demonstrations vs. Without demonstrations

Different kinds of fine tuning

Conventional fine tuning, classifying at hidden states of [CLS].
- Prompt-based fine tuning, predicting the best suited missing word.
- Prompt-based fine tuning with demonstrations to help model to know what is “great” / “terrible”.

Random templates/demonstrations vs.

Comparing our random templates/demonstrations as data augmentation to Synonym Replacement (SR), Random Insertion (RI), Random Swapping (RS), Random Deletion (RD) and EDA [4].

Overview of our proposed method.

Contrastive Learning for Prompt-based Few-shot Language Learners

Yiren Jian, Chongyang Gao, Soroush Vosoughi

Algorithm

With demonstrations vs. Without demonstrations

Different kinds of fine tuning

Conventional fine tuning, classifying at hidden states of [CLS].
- Prompt-based fine tuning, predicting the best suited missing word.
- Prompt-based fine tuning with demonstrations to help model to know what is “great” / “terrible”.

Random templates/demonstrations vs.

Comparing our random templates/demonstrations as data augmentation to Synonym Replacement (SR), Random Insertion (RI), Random Swapping (RS), Random Deletion (RD) and EDA [4].

Overview of our proposed method.

Contrastive Learning for Prompt-based Few-shot Language Learners

Yiren Jian, Chongyang Gao, Soroush Vosoughi

Algorithm

With demonstrations vs. Without demonstrations

Different kinds of fine tuning

Conventional fine tuning, classifying at hidden states of [CLS].
- Prompt-based fine tuning, predicting the best suited missing word.
- Prompt-based fine tuning with demonstrations to help model to know what is “great” / “terrible”.

Random templates/demonstrations vs.

Comparing our random templates/demonstrations as data augmentation to Synonym Replacement (SR), Random Insertion (RI), Random Swapping (RS), Random Deletion (RD) and EDA [4].

Overview of our proposed method.

Contrastive Learning for Prompt-based Few-shot Language Learners

Yiren Jian, Chongyang Gao, Soroush Vosoughi

Algorithm

With demonstrations vs. Without demonstrations

Different kinds of fine tuning

Conventional fine tuning, classifying at hidden states of [CLS].
- Prompt-based fine tuning, predicting the best suited missing word.
- Prompt-based fine tuning with demonstrations to help model to know what is “great” / “terrible”.

Random templates/demonstrations vs.

Comparing our random templates/demonstrations as data augmentation to Synonym Replacement (SR), Random Insertion (RI), Random Swapping (RS), Random Deletion (RD) and EDA [4].

Overview of our proposed method.