Motivations

• Extend prototypical networks
• Incorporate unlabeled data
Motivations

- Extend prototypical networks
- Incorporate unlabeled data
- Use data augmentation for more robustness
Motivations

- Extend prototypical networks
- Incorporate unlabeled data
- Use data augmentation for more robustness
- Data-augmented texts should be as diverse as possible
Back-translation [7] is a poor data augmentation strategy
Back-translation [7] is a poor data augmentation strategy

→ We use BART [13] for Natural Language Generation
Back-translation [7] is a poor data augmentation strategy

→ We use BART [13] for Natural Language Generation

We train the pre-trained BART model on a paraphrasing task, using the following datasets:

- Quora [21]
- Google’s PAWS [31]
- MSR Paraphrase corpus [32]
Constrained generation for more diversity

Diverse Beam Search (DBS [27]), an extension of the regular beam search decoding strategy, generates diverse sentences.

Figure 1: Comparing image captioning outputs decoded by BS (top) and our method, Diverse Beam Search (middle) – we notice that BS captions are near-duplicates with similar shared paths in the search tree and minor variations in the end. In contrast, DBS captions are significantly diverse and similar to the variability in human-generated ground truth captions (bottom).
Constrained generation for more diversity

We further constraint the paraphrase generation using:

- **uni-gram masking**: some tokens (according to a probability $p_{mask}$) of the input sentence cannot appear in the generated paraphrase.

- **bi-gram masking**: all bi-grams in the input sentence cannot appear in the generated paraphrase.
Paraphrase examples

<table>
<thead>
<tr>
<th>orig:</th>
<th>How long will my transfer be pending for?</th>
</tr>
</thead>
<tbody>
<tr>
<td>back:</td>
<td>How long will my transfer be on hold?</td>
</tr>
<tr>
<td>dbs:</td>
<td>How long will my transfer be pending? I am in first year.</td>
</tr>
<tr>
<td>dbs_unigram:</td>
<td>When are all transfers coming up and how many days are they expected?</td>
</tr>
<tr>
<td>dbs_bigram:</td>
<td>If I have a transfer for a while, how long should I wait for it?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>orig:</th>
<th>I am not sure where my phone is.</th>
</tr>
</thead>
<tbody>
<tr>
<td>back:</td>
<td>I don’t know where my phone is.</td>
</tr>
<tr>
<td>dbs:</td>
<td>I am not really sure where my phone is located</td>
</tr>
<tr>
<td>dbs_unigram:</td>
<td>How can I find the location of any Android mobile</td>
</tr>
<tr>
<td>dbs_bigram:</td>
<td>I don’t know where is my cell phone</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>orig:</th>
<th>can you play m3 file</th>
</tr>
</thead>
<tbody>
<tr>
<td>back:</td>
<td>can you read m3 file</td>
</tr>
<tr>
<td>dbs:</td>
<td>M3 files: can I play the entire M3 file?</td>
</tr>
<tr>
<td>dbs_unigram:</td>
<td>Is there any way to play 3M files on Earth without downloading it</td>
</tr>
<tr>
<td>dbs_bigram:</td>
<td>Is there any way to play M3 files on Windows?</td>
</tr>
</tbody>
</table>

**Table 5:** Examples of sentences paraphrased using back translation, vanilla diverse beam search, DBS with unigram masking, and DBS with bigram masking
Figure 12: Step 0: Pick unlabeled data
Figure 13: Step 1: Generate paraphrases
Figure 14: Step 2: Compute the prototypes, as average of paraphrases
Unsupervised Loss

Figure 15: Step 3: Compare unlabeled data to prototypes
**PROTAUGMENT’s framework**

**Figure 16:** The PROTAUGMENT framework
Results

Datasets Accuracy stats

<table>
<thead>
<tr>
<th>Data Profile</th>
<th>Method</th>
<th>Banking</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>K = 1</td>
<td>K = 5</td>
<td>K = 1</td>
<td>K = 5</td>
<td>K = 1</td>
<td>K = 5</td>
<td>K = 1</td>
<td>K = 5</td>
<td>K = 1</td>
<td>K = 5</td>
</tr>
<tr>
<td>low profile</td>
<td>Prototypical Network</td>
<td>82.20</td>
<td>91.57</td>
<td>74.37</td>
<td>86.48</td>
<td>80.06</td>
<td>89.62</td>
<td>94.29</td>
<td>98.10</td>
<td>82.73 ± 2.32</td>
<td>91.44 ± 1.92</td>
</tr>
<tr>
<td></td>
<td>ours w/ BT</td>
<td>83.83</td>
<td>92.16</td>
<td>78.70</td>
<td>89.36</td>
<td>80.84</td>
<td>90.87</td>
<td>94.06</td>
<td>97.62</td>
<td>84.36 ± 1.15</td>
<td>92.50 ± 0.94</td>
</tr>
<tr>
<td></td>
<td>ours w/ DBS</td>
<td>83.10</td>
<td>92.56</td>
<td>80.06</td>
<td>90.21</td>
<td>82.31</td>
<td>91.64</td>
<td>93.70</td>
<td>97.83</td>
<td>84.80 ± 1.26</td>
<td>93.06 ± 0.99</td>
</tr>
<tr>
<td></td>
<td>ours w/ DBS+bigram</td>
<td>86.04</td>
<td>93.55</td>
<td>82.09</td>
<td>91.57</td>
<td>83.60</td>
<td>92.71</td>
<td>95.11</td>
<td>98.23</td>
<td>86.71 ± 1.14</td>
<td>94.01 ± 1.05</td>
</tr>
<tr>
<td></td>
<td>ours w/ DBS+unigram</td>
<td>87.23</td>
<td>94.29</td>
<td>83.70</td>
<td>91.29</td>
<td>85.16</td>
<td>93.00</td>
<td>95.92</td>
<td>98.56</td>
<td>88.00 ± 1.22</td>
<td>94.29 ± 0.76</td>
</tr>
<tr>
<td>full profile</td>
<td>Prototypical Network</td>
<td>86.28</td>
<td>93.94</td>
<td>77.09</td>
<td>89.02</td>
<td>82.76</td>
<td>91.37</td>
<td>96.05</td>
<td>98.61</td>
<td>85.55 ± 2.20</td>
<td>93.24 ± 1.22</td>
</tr>
<tr>
<td></td>
<td>ours w/ BT</td>
<td>87.64</td>
<td>94.47</td>
<td>81.31</td>
<td>91.44</td>
<td>84.14</td>
<td>92.67</td>
<td>95.19</td>
<td>98.36</td>
<td>87.02 ± 1.36</td>
<td>94.23 ± 0.82</td>
</tr>
<tr>
<td></td>
<td>ours w/ DBS</td>
<td>86.94</td>
<td>94.50</td>
<td>82.35</td>
<td>91.68</td>
<td>84.42</td>
<td>92.62</td>
<td>94.85</td>
<td>98.41</td>
<td>87.14 ± 1.36</td>
<td>94.30 ± 0.60</td>
</tr>
<tr>
<td></td>
<td>ours w/ DBS+bigram</td>
<td>88.14</td>
<td>94.70</td>
<td>84.05</td>
<td>92.14</td>
<td>85.29</td>
<td>93.23</td>
<td>95.77</td>
<td>98.50</td>
<td>88.31 ± 1.43</td>
<td>94.64 ± 0.59</td>
</tr>
<tr>
<td></td>
<td>ours w/ DBS+unigram</td>
<td>89.56</td>
<td>94.71</td>
<td>84.34</td>
<td>92.55</td>
<td>86.11</td>
<td>93.70</td>
<td>96.49</td>
<td>98.74</td>
<td>89.13 ± 1.13</td>
<td>94.92 ± 0.57</td>
</tr>
</tbody>
</table>

Table 6: 5-way 1-shots and 5-way 5-shots accuracy on the test sets for each dataset. For each data profile, we highlight the best method in **bold**. We underline the methods on the low profile which perform better than the Prototypical Networks on the full profile.
Figure 17: Performance when changing the $p_{mask}$ value

- In terms of masking, **more** is better than **less**
- The optimal value for $p_{mask}$ is around 0.7
Key Findings

- PROT AUGMENT vastly out-performs prototypical networks
Key Findings

- PROTAUGMENT vastly out-performs prototypical networks
- Using constrained paraphrase generation is way better than using back-translated sentences
Key Findings

- PROT AUGMENT vastly out-performs prototypical networks
- Using constrained paraphrase generation is way better than using back-translated sentences
- PROT AUGMENT also yields large increments even using the full data profile
Key Findings

- PROTAUGMENT vastly out-performs prototypical networks
- Using constrained paraphrase generation is way better than using back-translated sentences
- PROTAUGMENT also yields large increments even using the full data profile

This word was (very) recently accepted as long paper at ACL2021:
Future Work
Future work

Test this unsupervised loss on other meta-learning models

Extend paraphrase generation in the multilingual case
Feel free to ask questions!

contact: thomas.dopierre@protonmail.com

code: https://github.com/tdopierre
References


[27] Ashwin K. Vijayakumar, Michael Cogswell, Ramprasaath R. Selvaraju, Qing Sun, Stefan Lee, David J. Crandall, and Dhruv Batra. Diverse beam search for improved description of complex scenes. In Sheila A. McIlraith and Kilian Q. Weinberger, editors, Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of
Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018, pages 7371–7379. AAAI Press, 2018.


