Curriculum-Meta Learning for Order-Robust Continual Relation Extraction

Authors:
Tongtong Wu, Xuekai Li, Yuan-Fang Li, Reza Haffari, Guilin Qi,
Yujin Zhu, Guoqiang Xu
Repoter: Xuekai Li
Part 1 Introduction
1.1 Background

Relation Extraction

unstructured text → structured facts triples

WIKIDATA
1.1 Background

Continual Relation Extraction

Task 1

Task 2

Task 3

... 

Task n

Different Relations

Support Set

Valid Set

Test Set

2020-12-18
1.2 Challenge

Catastrophic Forgetting

Continual Learning

Train

Valid

Task 1

Task 2

Task 3
1.2 Challenge

Catastrophic Forgetting

When a neural network is utilized to learn a sequence of tasks, the learning of the later tasks may degrade the performance of the learned model for the previous tasks.
1.2 Challenge

Order-sensitivity

The performance of the tasks vary based on the order of the task arrival sequence.

- CF incurred by the different sequences of previous tasks
- The unidirectional knowledge transfer from the previous tasks.

(i) Ethical AI considerations in continual learning, e.g. fairness in the medical domain [Yoon et al. 2020];
(ii) Benchmarking of continual learning algorithms as most of the existing works pick an arbitrary and random sequence of the given tasks for the evaluation [Chen et al. 2018];
(iii) Uncertainty to the quality of extracted knowledge in the realistic scenario for knowledge base population, where the model is faced with only one sequence.
1.3 Our Contribution

Three Contributions:

• A novel curriculum-based continual learning method tackling the order-sensitivity and catastrophic forgetting problems in continual relation extraction.

• A new relation representation learning method via the conceptual distribution of domain and range of relations.

• Comprehensive experiments to analyze the order-sensitivity and catastrophic forgetting problems in state-of-the-art models.
The Next

Part 2

Related Works
2 Related Works

Continual Learning

GEM [Lopez-Paz et al. 2017]  Experience Replay-based Model
EWC [Kirkpatrick et al. 2016]  Weight Consolidation-based Model
  - R-EWC [Liu et al. 2016]

Continual Relation Extraction

EA-EMR [Wang et al. 2019]
MLLRE [Obamuyide et al. 2019]
EMAR [Han et al. 2020]
Part 3 Curriculum-meta Learning
3.1 Continual Relation Extraction

\[ l = \max \{0, y - \sin(r^+, x) + \sin(r^-, x)\} \]
3.2 Framework

Our Framework

EA - EMR
3.3 Curriculum-Meta Learning

Meta-Training

Conventional Machine Learning:

CML
3.3 Curriculum-Meta Learning

**Meta-Training**

Conventional Machine Learning:

Meta-learning:

\[ \theta_{t+1} = \theta + \epsilon \sum_{i=0}^{n} (\theta^*_i - \theta_t) \]
3.3 Curriculum-Meta Learning

Curriculum-based Memory Replay

1. Assessing the difficulty of tasks.
2. Sampling instances from the memory.
3. Ranking the sampled instances by a certain strategy, inducting the model to learn the bias between the current task and observed similar tasks.
3.4 Knowledge-based Curriculum

KB-C

Difficulty Function:

\[ DI_i : = \frac{1}{K-1} \sum_{j=1; j \neq i}^{K} S_i^j \]

\[ S_i^j = \frac{1}{M \times N} \sum_{m=1}^{M} \sum_{n=1}^{N} s_m^j \]
3.4 Knowledge-based Curriculum

KB-C

Representation Function:

\[
m(\phi, i_j) = \min \sum_{(h, t : r) \in i_j} \left[ -\log P_\phi(h'|r) - \log P_\phi(t'|r) \right]
\]
3.4 Knowledge-based Curriculum

KB-C

Sampling Strategy:
Select and sort memory-stored instances of the most similar relations to current task.

\[
\theta_{t+1} = \theta + \epsilon \frac{1}{n} \sum_{i=0}^{n} (\theta_i^* - \theta_t)
\]
Part 4
Experiments and Discussion
4.1 Experiment Settings

Baselines
Vanilla Model [Yu et al. 2017]
EA-EMR [Wang et al. 2019]
MLLRE [Obamuyide et al. 2019]
EMAR [Han et al. 2020]

Benchmarks [Wang et al. 2019]
Lifelong - Fewrel: 80 relations and 700 instances per relation.
Lifelong - SimpleQuestions: 1,785 relations and totally 72,238 instances.
Lifelong - Tacred: 42 relations and totally 21,784 instances.
4.2 Main Results

The average accuracy $\text{Acc}_a$ and whole accuracy $\text{Acc}_w$, with error bounds $\text{EB}$, on the test sets of observed tasks at the final time step.

Metrics:

\[
\text{Acc}_w = \text{acc}_{f,1_{\text{test}}} \\
\text{Acc}_a = \frac{1}{k} \sum_{i=1}^{k} \text{acc}_{f,1_{i_{\text{test}}}} \\
\text{EB} = Z_a \times \frac{\delta}{\sqrt{n}}
\]
### 4.3 Case Study

A case study of EA-EMR, MLLRE and Our CML on the FewRel dataset.

#### Summary:
CML alleviate the order-sensitivity problem.

#### EA-EMR

<table>
<thead>
<tr>
<th>taskID</th>
<th>$T_0$</th>
<th>$T_1$</th>
<th>$T_2$</th>
<th>$T_3$</th>
<th>$T_4$</th>
<th>$T_5$</th>
<th>$T_6$</th>
<th>$T_7$</th>
<th>$T_8$</th>
<th>$T_9$</th>
<th>$\text{Acc}_a$</th>
<th>$\text{Acc}_w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_0$</td>
<td>93.3</td>
<td>36.7</td>
<td>61.7</td>
<td>57.7</td>
<td>100.0</td>
<td>26.5</td>
<td>35.7</td>
<td>58.1</td>
<td>54.1</td>
<td>50.6</td>
<td>76.0</td>
<td>60.9</td>
</tr>
<tr>
<td>$P_1$</td>
<td>91.9</td>
<td>42.4</td>
<td>62.5</td>
<td>64.6</td>
<td>99.3</td>
<td>24.5</td>
<td>60.3</td>
<td>62.2</td>
<td>61.8</td>
<td>74.0</td>
<td>61.7</td>
<td></td>
</tr>
<tr>
<td>$P_2$</td>
<td>92.6</td>
<td>42.7</td>
<td>60.6</td>
<td>66.1</td>
<td>100.0</td>
<td>25.8</td>
<td>55.4</td>
<td>69.3</td>
<td>71.1</td>
<td>54.5</td>
<td>73.6</td>
<td>61.3</td>
</tr>
<tr>
<td>$P_3$</td>
<td>93.3</td>
<td>39.2</td>
<td>68.4</td>
<td>60.2</td>
<td>100.0</td>
<td>34.0</td>
<td>57.3</td>
<td>71.4</td>
<td>85.2</td>
<td>58.3</td>
<td>74.9</td>
<td>63.4</td>
</tr>
<tr>
<td>$P_4$</td>
<td>89.6</td>
<td>34.4</td>
<td>64.0</td>
<td>59.9</td>
<td>100.0</td>
<td>35.1</td>
<td>65.8</td>
<td>71.0</td>
<td>77.8</td>
<td>72.5</td>
<td>75.6</td>
<td>55.8</td>
</tr>
<tr>
<td>$P_5$</td>
<td>91.1</td>
<td>60.9</td>
<td>56.9</td>
<td>64.2</td>
<td>99.3</td>
<td>45.7</td>
<td>77.9</td>
<td>71.1</td>
<td>67.4</td>
<td>79.9</td>
<td>75.1</td>
<td>58.3</td>
</tr>
<tr>
<td>$P_6$</td>
<td>94.1</td>
<td>66.9</td>
<td>72.1</td>
<td>55.5</td>
<td>100.0</td>
<td>53.0</td>
<td>77.2</td>
<td>74.3</td>
<td>84.4</td>
<td>74.0</td>
<td>74.5</td>
<td>59.2</td>
</tr>
<tr>
<td>$P_7$</td>
<td>94.8</td>
<td>73.6</td>
<td>85.4</td>
<td>53.6</td>
<td>98.6</td>
<td>61.6</td>
<td>80.4</td>
<td>81.5</td>
<td>92.6</td>
<td>86.5</td>
<td>76.0</td>
<td>62.0</td>
</tr>
<tr>
<td>$P_8$</td>
<td>94.8</td>
<td>74.6</td>
<td>85.2</td>
<td>85.0</td>
<td>98.6</td>
<td>69.1</td>
<td>79.3</td>
<td>75.2</td>
<td>87.4</td>
<td>92.9</td>
<td>74.2</td>
<td>54.5</td>
</tr>
<tr>
<td>$P_9$</td>
<td>91.1</td>
<td>76.9</td>
<td>91.3</td>
<td>89.4</td>
<td>97.9</td>
<td>65.4</td>
<td>77.6</td>
<td>83.9</td>
<td>93.3</td>
<td>85.8</td>
<td>76.1</td>
<td>58.9</td>
</tr>
</tbody>
</table>

#### CML

<table>
<thead>
<tr>
<th>runID</th>
<th>$T_0$</th>
<th>$T_1$</th>
<th>$T_2$</th>
<th>$T_3$</th>
<th>$T_4$</th>
<th>$T_5$</th>
<th>$T_6$</th>
<th>$T_7$</th>
<th>$T_8$</th>
<th>$T_9$</th>
<th>$\text{Acc}_a$</th>
<th>$\text{Acc}_w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_0$</td>
<td>88.1</td>
<td>39.9</td>
<td>49.9</td>
<td>59.9</td>
<td>100.0</td>
<td>23.8</td>
<td>28.2</td>
<td>48.5</td>
<td>52.6</td>
<td>53.9</td>
<td>76.8</td>
<td>62.7</td>
</tr>
<tr>
<td>$P_1$</td>
<td>82.2</td>
<td>43.5</td>
<td>68.8</td>
<td>33.9</td>
<td>100.0</td>
<td>30.7</td>
<td>56.7</td>
<td>39.6</td>
<td>71.1</td>
<td>59</td>
<td>73.4</td>
<td>63.9</td>
</tr>
<tr>
<td>$P_2$</td>
<td>83</td>
<td>49.9</td>
<td>66.7</td>
<td>75.9</td>
<td>100</td>
<td>26.7</td>
<td>57.4</td>
<td>74.7</td>
<td>58.5</td>
<td>64.1</td>
<td>72.6</td>
<td>61.2</td>
</tr>
<tr>
<td>$P_3$</td>
<td>91.9</td>
<td>48.3</td>
<td>76.3</td>
<td>77</td>
<td>100</td>
<td>23.2</td>
<td>54.3</td>
<td>77</td>
<td>83.7</td>
<td>50.6</td>
<td>62.8</td>
<td>59</td>
</tr>
<tr>
<td>$P_4$</td>
<td>90.4</td>
<td>44.1</td>
<td>73.6</td>
<td>79.6</td>
<td>100</td>
<td>43.5</td>
<td>47.5</td>
<td>69.6</td>
<td>86.7</td>
<td>77.6</td>
<td>77.7</td>
<td>58.4</td>
</tr>
<tr>
<td>$P_5$</td>
<td>97</td>
<td>42.8</td>
<td>71.2</td>
<td>79.9</td>
<td>100</td>
<td>46</td>
<td>74.9</td>
<td>68.5</td>
<td>80</td>
<td>76.8</td>
<td>79.5</td>
<td>61.6</td>
</tr>
<tr>
<td>$P_6$</td>
<td>98.5</td>
<td>79.2</td>
<td>69.3</td>
<td>75.5</td>
<td>100</td>
<td>52.4</td>
<td>82.6</td>
<td>88.1</td>
<td>88.1</td>
<td>80.4</td>
<td>74.6</td>
<td>58.7</td>
</tr>
<tr>
<td>$P_7$</td>
<td>97</td>
<td>80.2</td>
<td>92.1</td>
<td>67.9</td>
<td>100</td>
<td>57.1</td>
<td>81.9</td>
<td>87.6</td>
<td>93.3</td>
<td>81.4</td>
<td>67.4</td>
<td>56.3</td>
</tr>
<tr>
<td>$P_8$</td>
<td>91.9</td>
<td>81.5</td>
<td>91.9</td>
<td>92.7</td>
<td>99.3</td>
<td>68</td>
<td>86.1</td>
<td>93.4</td>
<td>92.6</td>
<td>90.8</td>
<td>77.4</td>
<td>56.1</td>
</tr>
<tr>
<td>$P_9$</td>
<td>90.0</td>
<td>89.7</td>
<td>97.4</td>
<td>96</td>
<td>100</td>
<td>82.1</td>
<td>91.9</td>
<td>94.7</td>
<td>99.3</td>
<td>93.1</td>
<td>77.7</td>
<td>58.7</td>
</tr>
</tbody>
</table>

#### Summary:

| $\mu$  | 92.7  | 54.8  | 70.8  | 65.6  | 99.4  | 44.1  | 66.7  | 71.8  | 77.6  | 71.7  | 75.0  |
| $\delta$ | 1.73  | 17.34 | 12.22 | 12.09 | 0.77  | 17.25 | 14.61 | 8.97  | 13.36 | 14.75 | 0.91  | 2.83   |

2020-12-18
## 4.3 Case Study

A case study of EA-EMR, MLLRE and Our CML on the FewRel dataset.

### CML

<table>
<thead>
<tr>
<th>taskID</th>
<th>$T_0$</th>
<th>$T_1$</th>
<th>$T_2$</th>
<th>$T_3$</th>
<th>$T_4$</th>
<th>$T_5$</th>
<th>$T_6$</th>
<th>$T_7$</th>
<th>$T_8$</th>
<th>$T_9$</th>
<th>Acc$_a$</th>
<th>Acc$_w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_0$</td>
<td>93.3</td>
<td>36.7</td>
<td>61.7</td>
<td>57.7</td>
<td>100.0</td>
<td>26.5</td>
<td>35.7</td>
<td>58.1</td>
<td>54.1</td>
<td>50.6</td>
<td>76.0</td>
<td>60.9</td>
</tr>
<tr>
<td>$P_1$</td>
<td>91.9</td>
<td>42.4</td>
<td>62.5</td>
<td>64.6</td>
<td>99.3</td>
<td>24.5</td>
<td>60.3</td>
<td>56.2</td>
<td>62.2</td>
<td>61.8</td>
<td>74.0</td>
<td>61.7</td>
</tr>
<tr>
<td>$P_2$</td>
<td>92.6</td>
<td>42.7</td>
<td>60.6</td>
<td>66.1</td>
<td>100.0</td>
<td>25.8</td>
<td>55.4</td>
<td>69.3</td>
<td>71.1</td>
<td>54.5</td>
<td>73.6</td>
<td>61.3</td>
</tr>
<tr>
<td>$P_3$</td>
<td>93.3</td>
<td>39.2</td>
<td>68.4</td>
<td>60.2</td>
<td>100.0</td>
<td>34.0</td>
<td>57.3</td>
<td>71.4</td>
<td>85.2</td>
<td>58.3</td>
<td>74.9</td>
<td>63.4</td>
</tr>
<tr>
<td>$P_4$</td>
<td>89.6</td>
<td>34.4</td>
<td>64.0</td>
<td>59.9</td>
<td>100.0</td>
<td>35.1</td>
<td>65.8</td>
<td>71.0</td>
<td>77.8</td>
<td>72.5</td>
<td>75.6</td>
<td>55.8</td>
</tr>
<tr>
<td>$P_5$</td>
<td>91.1</td>
<td>60.9</td>
<td>56.9</td>
<td>64.2</td>
<td>99.3</td>
<td>45.7</td>
<td>77.9</td>
<td>77.1</td>
<td>67.4</td>
<td>79.9</td>
<td>75.1</td>
<td>58.3</td>
</tr>
<tr>
<td>$P_6$</td>
<td>94.1</td>
<td>66.9</td>
<td>72.1</td>
<td>55.5</td>
<td>100.0</td>
<td>53.0</td>
<td>77.2</td>
<td>74.3</td>
<td>84.4</td>
<td>74.0</td>
<td>74.5</td>
<td>59.2</td>
</tr>
</tbody>
</table>

### MLLRE

<table>
<thead>
<tr>
<th>taskID</th>
<th>$T_0$</th>
<th>$T_1$</th>
<th>$T_6$</th>
<th>$T_3$</th>
<th>$T_4$</th>
<th>$T_5$</th>
<th>$T_6$</th>
<th>$T_7$</th>
<th>$T_8$</th>
<th>$T_9$</th>
<th>Acc$_a$</th>
<th>Acc$_w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_0$</td>
<td>83.0</td>
<td>44.1</td>
<td>48.8</td>
<td>53.6</td>
<td>100.0</td>
<td>25.9</td>
<td>49.8</td>
<td>52.7</td>
<td>63.0</td>
<td>54.7</td>
<td>74.8</td>
<td>59.7</td>
</tr>
<tr>
<td>$P_1$</td>
<td>91.9</td>
<td>39.0</td>
<td>51.0</td>
<td>62.4</td>
<td>100.0</td>
<td>25.6</td>
<td>61.4</td>
<td>60.7</td>
<td>48.1</td>
<td>53.9</td>
<td>75.3</td>
<td>60.3</td>
</tr>
<tr>
<td>$P_2$</td>
<td>94.1</td>
<td>43.5</td>
<td>62.3</td>
<td>70.4</td>
<td>100.0</td>
<td>23.8</td>
<td>56.2</td>
<td>70.0</td>
<td>72.6</td>
<td>44.0</td>
<td>71.0</td>
<td>60.9</td>
</tr>
<tr>
<td>$P_3$</td>
<td>84.4</td>
<td>49.7</td>
<td>72.8</td>
<td>62.4</td>
<td>100.0</td>
<td>31.1</td>
<td>58.9</td>
<td>71.5</td>
<td>80.0</td>
<td>62.6</td>
<td>72.4</td>
<td>62.6</td>
</tr>
<tr>
<td>$P_4$</td>
<td>97.0</td>
<td>46.1</td>
<td>62.3</td>
<td>80.7</td>
<td>100.0</td>
<td>31.7</td>
<td>63.2</td>
<td>68.0</td>
<td>70.4</td>
<td>78.1</td>
<td>75.5</td>
<td>56.5</td>
</tr>
<tr>
<td>$P_5$</td>
<td>95.6</td>
<td>53.3</td>
<td>59.5</td>
<td>75.5</td>
<td>100.0</td>
<td>45.2</td>
<td>68.0</td>
<td>80.4</td>
<td>70.4</td>
<td>72.5</td>
<td>76.1</td>
<td>57.7</td>
</tr>
<tr>
<td>$P_6$</td>
<td>95.6</td>
<td>72.0</td>
<td>71.2</td>
<td>59.9</td>
<td>99.3</td>
<td>45.3</td>
<td>77.7</td>
<td>77.2</td>
<td>84.4</td>
<td>70.7</td>
<td>74.7</td>
<td>57.6</td>
</tr>
</tbody>
</table>

### Summary:

CML do alleviate the order-sensitivity problem.

---

2020-12-18
### Memory-Training Rate

<table>
<thead>
<tr>
<th></th>
<th>Train</th>
<th>100</th>
<th>200</th>
<th>all</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory</td>
<td>25</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>EA-EMR</td>
<td>$Acc_a$</td>
<td>70.7</td>
<td>75.5</td>
<td>74.8</td>
</tr>
<tr>
<td></td>
<td>$Acc_w$</td>
<td>53.2</td>
<td>57.4</td>
<td>59.8</td>
</tr>
<tr>
<td>MLLRE</td>
<td>$Acc_a$</td>
<td>68.4</td>
<td>72.1</td>
<td>70.2</td>
</tr>
<tr>
<td></td>
<td>$Acc_w$</td>
<td>51.9</td>
<td>57.8</td>
<td>56.8</td>
</tr>
<tr>
<td>EMAR</td>
<td>$Acc_a$</td>
<td>60.1</td>
<td>66.7</td>
<td>68.6</td>
</tr>
<tr>
<td></td>
<td>$Acc_w$</td>
<td>43.7</td>
<td>51.2</td>
<td>53.8</td>
</tr>
<tr>
<td>CML</td>
<td>$Acc_a$</td>
<td><strong>73.6</strong></td>
<td><strong>76.4</strong></td>
<td><strong>76.0</strong></td>
</tr>
<tr>
<td></td>
<td>$Acc_w$</td>
<td><strong>54.7</strong></td>
<td><strong>60.3</strong></td>
<td><strong>60.2</strong></td>
</tr>
</tbody>
</table>

**Summary:** The performance of CML is much better than MLLRE.
4.5 The effectiveness of KB-C

A New Metric

Average Forgetting Rate:

\[ Fr_{ag}^j := \frac{1}{K-1} \sum_{i=1}^{K-1} \frac{ac_{i+1} - ac_i}{ac_i} \]

\[ ac_i := \frac{1}{(J-1)!} \sum_{\pi \in \Pi_{[1,...,J]}} acc_i(\pi) \]

Average Forgetting Rate \( Fr_{ag} \) is used to evaluate the actual difficulty of each task based on the final result.
4.5 The effectiveness of KB-C

Prior Difficulty: \[ D_{\text{prior}}^i : = \frac{1}{K-1} \sum_{j=1; j \neq i}^{K} S_{ij} \]

Posterior Difficulty: \[ D_{\text{post}}^i : = Fr_{\text{ag}}^i \]

Summary: domain/range-based relation similarity is positively related to the difficulty of the task; Our Kb-C module reduces the interference of similar tasks.
The Next

Part 5

Conclusion
A Task: Continual Relation Extraction;

Two Problems: Catastrophic Forgetting & Order-Sensitivity;

Two Factors: Over-Fitting & Task Similarity;

Two Main Contribution: CML Framework and KB-C Module;

Three Future Works.
References

THANKS

2020.12.19