Unsupervised Model Personalization while Preserving Privacy and Scalability: An Open Problem

Matthias De Lange\textsuperscript{1}  
Ales Leonardis\textsuperscript{2}  
Xu Jia\textsuperscript{2}  
Gregory Slabaugh\textsuperscript{2}  
Sarah Parisot\textsuperscript{2,3}  
Tinne Tuytelaars\textsuperscript{1}

\textsuperscript{1}KU Leuven \quad \textsuperscript{2}Huawei, Noah’s Ark Lab \quad \textsuperscript{3}Mila

\{firstname.lastname\}@kuleuven.be \quad \{firstname.lastname\}@huawei.com

CVPR 2020
Roadmap

• Model Personalization: What and how?
• Novel Benchmarks
• Adaptation on the server
• Adapting locally
• Conclusion


Questions? matthias.delange@kuleuven.be

Questions? matthias.delange@kuleuven.be
Model Personalization

2 Ways

- User-Adaptation on the server:
  + High Capacity
Privacy


Questions? matthias.delange@kuleuven.be
Scalability

Thousands

USER

SERVER
Model Personalization
2 Ways

- User-Adaptation local on user device:
  - Low capacity
  + No privacy issues
  + No scalability issues

Questions? matthias.delange@kuleuven.be
Model Personalization
2 Ways

On the server:

+ High Capacity
- Privacy
- Scalability
- Unlabeled user-data

Locally:

- Low Capacity
+ Privacy
+ Scalability
- Unlabeled user-data
Model Personalization
2 Ways

On the server:
- High Capacity
- Privacy
- Scalability
- Unlabeled user-data

Locally:
- Low Capacity
- Privacy
- Scalability
- Unlabeled user-data

One Framework
Model Personalization

2 Ways

On the server:

+ High Capacity  
- Privacy  
- Scalability  
- Unlabeled user-data

Locally:

- Low Capacity  
+ Privacy  
+ Scalability  
- Unlabeled user-data

One Framework

2x adaptation
Dual User-Adaptation framework (DUA)
Roadmap

• Model Personalization: What and how?
• **Novel Benchmarks**
• Adaptation on the server
• Adapting locally
• Conclusion
Benchmarks

- 2 components
  - Users with different preferences (prior) → Validation/ Evaluation sets
  - Server with large dataset → Training set

- Task incremental continual learning, see [1]
  - Divide into sequence of tasks

Benchmarks: Numbers

5 tasks of 2 subsequent numbers

User 1

Test data
MNIST

Train data
MNIST

SERVER

User 2

Test data
SVHN

Train data
SVHN


Questions? matthias.delange@kuleuven.be
Benchmarks: CatPrior

Each user has 3 local category (scene) preferences per Task (Supercategory)


Questions? matthias.delange@kuleuven.be
Benchmarks: TransPrior

10 users, Each user has 1 transformation

Spatters, elastic transformation, saturation, defocus blur, Gaussian noise, brightness, Gaussian blur, jpeg compression, contrast and impulse noise.
Roadmap

• Model Personalization: What and how?
• Novel Benchmarks
• Adaptation on the server
• Adapting locally
• Conclusion


Questions? matthias.delange@kuleuven.be
Server adaptation in a fashion?
Server adaptation in a

Scalable

Privacy Preserving

Unsupervised

fashion?
Continual Learning

- Continual learning major focus on *Catastrophic Forgetting*
- Many of its properties suit our setting:

**Local scalability** $\rightarrow$ limit user resources to model learning multiple tasks

**Distributed scalability** $\rightarrow$ limit server resources to many personalized models e.g. task incremental with IMM [2]

Continual Learning

**Distributed scalability** with IMM

1. Learn task-specific models
2. Get model importance weights
3. Merge using importance weights

→ #models = #tasks


Questions? matthias.delange@kuleuven.be
Dual User-Adaptation framework (DUA)

USER

DUA

SERVER

USER
(1) Serverside Adaptation

USER

SERVER

✓ Scalable

TASK-SPECIFIC
(1) Serverside Adaptation

✓ Scalable
✓ Privacy
(1) Serverside Adaptation

✓ Scalable
✓ Privacy

Questions? matthias.delange@kuleuven.be
Merging IMM-models

- Originally importance weights diagonal Fisher Information Matrix (FIM)
  - Loss-based → Requires labels

- Instead unsupervised MAS [3] importance weights, based on output function?

<table>
<thead>
<tr>
<th>Data Setup</th>
<th>Model</th>
<th>MAS-IMM</th>
<th>FIM-IMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CatPrior</td>
<td>AlexNet</td>
<td>67.39 (0.73)</td>
<td>67.42 (0.23)</td>
</tr>
<tr>
<td></td>
<td>VGG11</td>
<td>76.77 (0.30)</td>
<td>76.29 (0.43)</td>
</tr>
<tr>
<td>TransPrior</td>
<td>AlexNet</td>
<td>46.51 (-0.14)</td>
<td>46.68 (-0.35)</td>
</tr>
<tr>
<td></td>
<td>VGG11</td>
<td>53.49 (-0.17)</td>
<td>53.14 (0.07)</td>
</tr>
<tr>
<td>Numbers</td>
<td>MLP</td>
<td>84.36 (-0.40)</td>
<td>87.68 (0.07)</td>
</tr>
</tbody>
</table>

Remote Adaptive Continual Learning (RACL)

✓ Scalable
✓ Privacy
✓ Unsupervised
User-Adaptation

- RACL/IMM $\rightarrow$ User-specific/General model
- MAS/FIM $\rightarrow$ Unsupervised/supervised importance weights

<table>
<thead>
<tr>
<th>Method</th>
<th>Alexnet</th>
<th>VGG11</th>
<th>MLP</th>
<th>Adapt.</th>
<th>Unsup.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CatPrior</td>
<td>TransPrior</td>
<td>CatPrior</td>
<td>TransPrior</td>
<td>Numbers</td>
</tr>
<tr>
<td>MAS-RACL</td>
<td>66.97 (0.88)</td>
<td>47.04 (-0.27)</td>
<td>77.32 (0.77)</td>
<td>53.59 (-0.14)</td>
<td>84.01 (-0.22)</td>
</tr>
<tr>
<td>MAS-IMM</td>
<td>67.39 (0.73)</td>
<td>46.51 (-0.14)</td>
<td>76.77 (0.30)</td>
<td>53.49 (-0.17)</td>
<td>84.36 (-0.40)</td>
</tr>
<tr>
<td>FIM-RACL</td>
<td>67.20 (0.73)</td>
<td>47.32 (-0.51)</td>
<td>76.53 (0.68)</td>
<td>53.73 (-0.13)</td>
<td>87.83 (0.30)</td>
</tr>
<tr>
<td>FIM-IMM</td>
<td>67.42 (0.23)</td>
<td>46.68 (-0.35)</td>
<td>76.29 (0.43)</td>
<td>53.14 (0.07)</td>
<td>87.68 (0.07)</td>
</tr>
</tbody>
</table>

Improvements insignificant $\rightarrow$ Why?
Analyzing Importance Weights

We found consistently:

1. High correlations for different datasets on a same model
2. Low correlation for same data on different models

*Importance weights indicate parameter importance for the specific model, rather than the data they are estimated from!*
Roadmap

- Model Personalization: What and how?
- Novel Benchmarks
- Adaptation on the server
- **Adapting locally**
- Conclusion
(1) Serverside Adaptation

✓ Scalable
✓ Privacy
✓ Unsupervised
(2) Local Adaptation

✓ Scalable
✓ Privacy
✓ Unsupervised
✓ 2 x Adaptation
Local Adaptation

- Adapt from general Server domain → User domain
- With limited user resources
  - Adapt Batch Normalization stats (AdaBN) → Unsupervised
  - Train few BN parameters (AdaBN-S) → Supervised

---

<table>
<thead>
<tr>
<th>Method</th>
<th>CatPrior</th>
<th>TransPrior</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BN</td>
<td>AdaBN</td>
</tr>
<tr>
<td>MAS-RACL</td>
<td>58.05 (2.74)</td>
<td>58.30 (2.34)</td>
</tr>
<tr>
<td>FIM-RACL</td>
<td>59.58 (2.14)</td>
<td>59.71 (1.61)</td>
</tr>
<tr>
<td>Task Experts</td>
<td>80.78 (5.61)</td>
<td>n/a</td>
</tr>
<tr>
<td>MAS-IMM</td>
<td>55.55 (2.69)</td>
<td>55.89 (2.69)</td>
</tr>
<tr>
<td>FIM-IMM</td>
<td>61.50 (-0.03)</td>
<td>61.35 (-0.46)</td>
</tr>
<tr>
<td>MAS</td>
<td>65.58 (3.96)</td>
<td>64.15 (4.04)</td>
</tr>
<tr>
<td>EWC</td>
<td>66.20 (2.88)</td>
<td>64.03 (3.43)</td>
</tr>
<tr>
<td>LWF</td>
<td>70.76 (0.73)</td>
<td>70.37 (0.43)</td>
</tr>
<tr>
<td>Joint</td>
<td>75.75 (n/a)</td>
<td>72.13 (n/a)</td>
</tr>
</tbody>
</table>
Local Adaptation

- Still open problem lightweight, unsupervised domain adaptation
- Relaxing unsupervised user training $\rightarrow$ Consistent improvements $\approx 3\%$

<table>
<thead>
<tr>
<th>Method</th>
<th>CatPrior</th>
<th>TransPrior</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BN</td>
<td>AdaBN</td>
</tr>
<tr>
<td>MAS-RACL</td>
<td>58.05 (2.74)</td>
<td>58.30 (2.34)</td>
</tr>
<tr>
<td>FIM-RACL</td>
<td>59.58 (2.14)</td>
<td>59.71 (1.61)</td>
</tr>
<tr>
<td>Task Experts</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>MAS-IMM</td>
<td>55.55 (2.69)</td>
<td>55.89 (2.69)</td>
</tr>
<tr>
<td>FIM-IMM</td>
<td>61.50 (-0.03)</td>
<td>61.35 (-0.46)</td>
</tr>
<tr>
<td>MAS</td>
<td>65.58 (3.96)</td>
<td>64.15 (4.04)</td>
</tr>
<tr>
<td>EWC</td>
<td>66.20 (2.88)</td>
<td>64.03 (3.43)</td>
</tr>
<tr>
<td>LWF</td>
<td>70.76 (0.73)</td>
<td>70.37 (0.43)</td>
</tr>
<tr>
<td>Joint</td>
<td>75.75 (n/a)</td>
<td>72.13 (n/a)</td>
</tr>
</tbody>
</table>
Roadmap

- Model Personalization: What and how?
- Novel Benchmarks
- Adaptation on the server
- Adapting locally
- Conclusion
Dual User-Adaptation

✓ Scalable
✓ Privacy
✓ Unsupervised
(2) Local Adaptation

✓ Scalable
✓ Privacy
✓ Unsupervised
✓ 2 x Adaptation
(2) Local Adaptation

- Scalable
- Privacy
- Unsupervised
- 2 x Adaptation

Open Problem:
- Data-dependent importance weights
- Domain Adaptation
Code

https://github.com/mattdl/DUA

Questions?

matthias.delange@kuleuven.be