Artificial Intelligence
Advanced Topics in AI & ML

Learning Frameworks: Meta-Learning, Few-Shot Learning, Multi-Tasking, and Multi-Modality

Aleksandr Petiushko

ML Research

SOFIA UNIVERSITY

Content

Content

1. Multi-tasking
2. Few-shot learning

Content

1. Multi-tasking
2. Few-shot learning
3. Meta-learning

Content

Motivation

- Common situation for a DL problem:
  - Have one modality (either text, or image, or video, or waveform)

Motivation

- Common situation for a DL problem:
  - Have one modality (either text, or image, or video, or waveform)
  - Have one task (e.g. recognition from a predefined set of classes)

Motivation

- Common situation for a DL problem:
  - Have one modality (either text, or image, or video, or waveform)
  - Have one task (e.g. recognition from a predefined set of classes)
  - Have a lot of data to train

Motivation

- Common situation for a DL problem:
  - Have one modality (either text, or image, or video, or waveform)
  - Have one task (e.g. recognition from a predefined set of classes)
  - Have a lot of data to train
- But sometimes:
  - Need to adapt quickly to the new task and small data

Motivation

- Common situation for a DL problem:
  - Have one modality (either text, or image, or video, or waveform)
  - Have one task (e.g. recognition from a predefined set of classes)
  - Have a lot of data to train
- But sometimes:
  - Need to adapt quickly to the new task and small data
  - Need to solve multiple tasks in parallel by a single model

Motivation

- Common situation for a DL problem:
  - Have one modality (either text, or image, or video, or waveform)
  - Have one task (e.g. recognition from a predefined set of classes)
  - Have a lot of data to train
- But sometimes:
  - Need to adapt quickly to the new task and small data
  - Need to solve multiple tasks in parallel by a single model
  - Need to work with completely different modalities

Motivation

- Common situation for a DL problem:
  - Have one modality (either text, or image, or video, or waveform)
  - Have one task (e.g. recognition from a predefined set of classes)
  - Have a lot of data to train
- But sometimes:
  - Need to adapt quickly to the new task and small data
  - Need to solve multiple tasks in parallel by a single model
  - Need to work with completely different modalities
- Direction to tackle these problems: **Representation Learning**

Multi-tasking

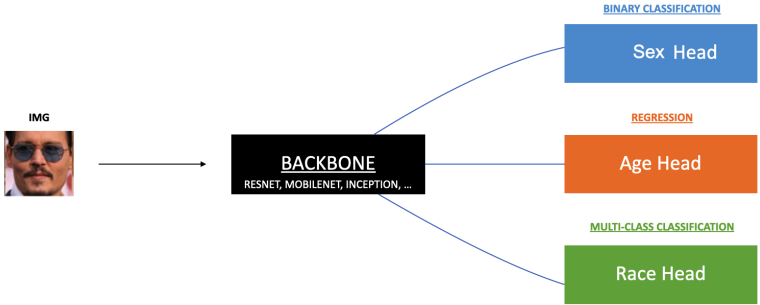
- How to solve different tasks by a single model, when we have enough training data?

Multi-tasking

- How to solve different tasks by a single model, when we have enough training data?
- It can be done simultaneously: we train the same feature extractor ("*backbone*") providing the common **representation** for all tasks with multiple heads on top solving these tasks; multiple objectives (losses) are optimized in parallel

Multi-tasking

- How to solve different tasks by a single model, when we have enough training data?
- It can be done simultaneously: we train the same feature extractor ("*backbone*") providing the common **representation** for all tasks with multiple heads on top solving these tasks; multiple objectives (losses) are optimized in parallel
- Sometimes different tasks can help each other (in data-greedy regime)
- Read material: link



Few-shot Learning and Meta-learning framework

- How to solve different tasks by a single model, when we don't have enough training data?

Few-shot Learning and Meta-learning framework

- How to solve different tasks by a single model, when we don't have enough training data?
- Few-shot learning problem setting:
  - A lot of training data but classes are not the same as during inference

Few-shot Learning and Meta-learning framework

- How to solve different tasks by a single model, when we don't have enough training data?
- Few-shot learning problem setting:
  - A lot of training data but classes are not the same as during inference
  - **N-way-K-shot classification** during the inference: we have a *support* set with N classes with K examples of each, and a *query* set of examples of unknown classes (which we need to infer)

Few-shot Learning and Meta-learning framework

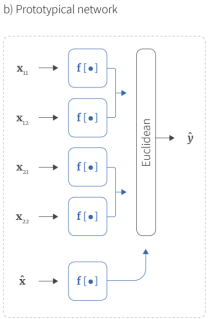
- How to solve different tasks by a single model, when we don't have enough training data?
- Few-shot learning problem setting:
  - A lot of training data but classes are not the same as during inference
  - **N-way-K-shot classification** during the inference: we have a *support* set with N classes with K examples of each, and a *query* set of examples of unknown classes (which we need to infer)
- The framework for it is called **Meta-learning** (for any task description): e.g., in case of few-shot learning we just need to sample multiple times different support sets with different classes
- Read material: link

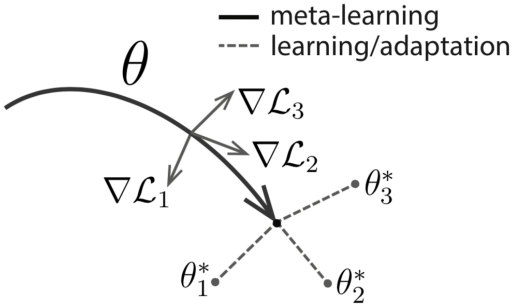
Meta-learning

- The main approaches to tackle the meta-learning problem:
  - Metric Learning (e.g., by the Prototypical Networks): find the representation template for support set classes and calculate the 1-NN based on the Euclidean distance between query set representation

Meta-learning

- The main approaches to tackle the meta-learning problem:
  - Metric Learning (e.g., by the Prototypical Networks): find the representation template for support set classes and calculate the 1-NN based on the Euclidean distance between query set representation
  - Gradient-based (e.g., by MAML): find the best NN weight initialization by averaging the gradients of losses (objectives) of multiple tasks
- Read material: link



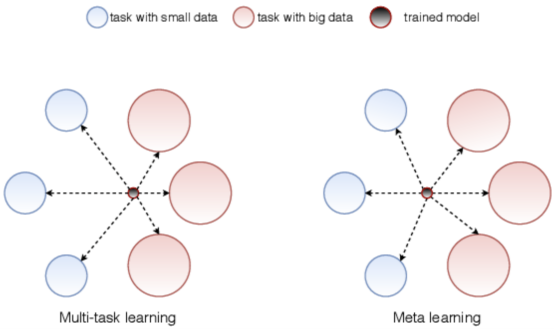


Meta-learning vs Multi-tasking

- If multi-task learning is done sequentially, then we come to meta-learning

Meta-learning vs Multi-tasking

- If multi-task learning is done sequentially, then we come to meta-learning
- If multi-task learning is done simultaneously, then we come to multi-tasking
- Read material: link



Multi-modality

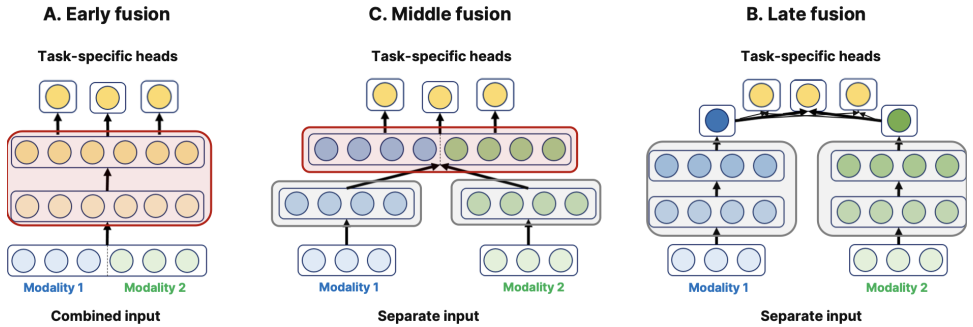
- It seems that incorporating different modalities (like human does: vision, hearing, flair, etc.) can improve the performance of an ML modal

Multi-modality

- It seems that incorporating different modalities (like human does: vision, hearing, flair, etc.) can improve the performance of an ML modal
- Two main modalities to combine now: text and images

Multi-modality

- It seems that incorporating different modalities (like human does: vision, hearing, flair, etc.) can improve the performance of an ML modal
- Two main modalities to combine now: text and images
- Techniques differ in the way where we fuse the modality: early or late fusion (remember: *representation learning*!)
- Read material: link



Takeaway notes

1. Read all the mentioned links

Takeaway notes

1. Read all the mentioned links
2. Meta-Learning is one of the hardest setting in DL

Takeaway notes

1. Read all the mentioned links
2. Meta-Learning is one of the hardest setting in DL
3. We can solve multiple tasks in parallel (multi-tasking) or sequentially (meta-learning)

Takeaway notes

1. Read all the mentioned links
2. Meta-Learning is one of the hardest setting in DL
3. We can solve multiple tasks in parallel (multi-tasking) or sequentially (meta-learning)
4. We can fuse multiple modalities in one Neural Net

Takeaway notes

1. Read all the mentioned links
2. Meta-Learning is one of the hardest setting in DL
3. We can solve multiple tasks in parallel (multi-tasking) or sequentially (meta-learning)
4. We can fuse multiple modalities in one Neural Net
5. A lot of approaches are based on the Representation Learning!

Thank you!