Domain Adaptation for Cross-Domain Studies in Astronomy:

Merging Galaxies Identification

Aleksandra Ćiprijanović
Research Associate
Scientific Computing Division
aleksand@fnal.gov

Where the Earth Meets the Sky
27-28 May 2021
Cosmic Dawn Center at DTU
1. Astro example and what I work on
2. What is domain discrepancy?
3. Domain adaptation - two methods
4. How does domain adaptation help?
Merging galaxies

**WHY**
To understand the evolution of our Universe (galaxy mergers lead to hierarchical formation of structures).

**HOW**
Leverage a large sample of merging galaxies to study.

**PROBLEMS**
Standard methods require knowledge about the morphology (we need for precise observations). Visual classification is very time consuming and prone to errors.

**SOLUTION**
Large simulations (we know the ground truth) + machine learning
Where are differences coming from?

Simulations are not perfect
- physics missing, computational resources

Dataset shift in astronomy

Making mock images is hard - adding noise, PSF, telescope imperfections

Use data from multiple telescopes with different specs

Simulation (source) LABELED!

Real (target) UNLABELED!
Two experiments

Simulation ⟷ Simulation + Noise

Simulation ⟷ Observations
Two experiments

- Illustris simulation
  - source (sim. + Hubble PSF)
  - target (sim. + PSF + random sky shot noise)
- Distant mergers at $z=2$
- 2233 individual galaxies
- $\sim 15,000$ images

---

Vogelsberger et al. (2014)

Čiprianović et al. (2020)
Total Loss = **Task Loss + Transfer Loss**

**Task loss** - very often categorical cross-entropy loss

$$L_{\text{cross-entropy}}(\hat{y}, y) = - \sum_i y_i \log(\hat{y}_i)$$

**Transfer loss** - domain alignment

**Maximum Mean Discrepancy**

Non-parametric distance between two probability distributions (distance of the mean embeddings of the samples in the kernel space).

**Adversarial training on domain labels**

Using Domain Adversarial Neural Network (DANN) to force domain-invariant feature extraction.
Maximum Mean Discrepancy - MMD

Are $P$ and $Q$ different?

Observe $X = \{x_1, \ldots, x_n\} \sim P$

Observe $Y = \{y_1, \ldots, y_n\} \sim Q$

$\hat{\mu}_P(v) := \frac{1}{m} \sum_{i=1}^{m} k(x_i, v)$

$\hat{\mu}_Q(v)$: mean embedding of $Q$

$\text{witness}(v) = \hat{\mu}_P(v) - \hat{\mu}_Q(v)$

$\text{MMD}^2 = ||\text{witness}(v)||_F^2$

$$= \frac{1}{n(n-1)} \sum_{i \neq j} k(x_i, x_j) + \frac{1}{n(n-1)} \sum_{i \neq j} k(y_i, y_j)$$

$$- \frac{2}{n^2} \sum_{i,j} k(x_i, y_j)$$

Minimize MMD loss by maximizing cross-similarities $\Rightarrow$ we find domain invariant features!

From Arthur Gretton (NIPS 2016 Workshop on Adversarial Learning, Barcelona Spain)
Domain Adversarial Neural Networks - DANNs

DANN - feature extractor + label predictor + domain classifier

- **Gradient reversal layer** - multiplies the gradient by a negative constant during the backpropagation.
- Results in the extraction of **domain-invariant features**.
- Only source domain images are labeled during training.

Ganin et al. (2016)
Results

DeepMerge II: Building Robust Deep Learning Algorithms for Merging Galaxy Identification Across Domains

arxiv:2103.01373

Grad-CAM (merger class)

What is the network focusing on?

<table>
<thead>
<tr>
<th>Source Domain</th>
<th>Target Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>noDA</td>
<td>85% 58%</td>
</tr>
<tr>
<td>MMD</td>
<td>87% 77%</td>
</tr>
<tr>
<td>DANN</td>
<td>87% 79%</td>
</tr>
</tbody>
</table>
Results

DeepMerge II: Building Robust Deep Learning Algorithms for Merging Galaxy Identification Across Domains
arxiv:2103.01373

Grad-CAM (merger class)

What is the network focusing on?

<table>
<thead>
<tr>
<th></th>
<th>Source Domain</th>
<th>Target Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>noDA</td>
<td>85%</td>
<td>58%</td>
</tr>
<tr>
<td>MMD</td>
<td>87%</td>
<td>77%</td>
</tr>
<tr>
<td>DANN</td>
<td>87%</td>
<td>79%</td>
</tr>
</tbody>
</table>
Two experiments

Simulation → Simulation + Noise

Simulation → Observations
Two experiments

- **Source**: Illustris nearby galaxies
  - $z=0$
  - small dataset (44 mergers)!

- **Target**: Real galaxies - SDSS:
  - small dataset (310 mergers)!
  - $z<0.1$
  - very different, only simple examples!
  - labeled by humans!

- $\sim 6000$ images

Vogelsberger et al. (2014)
Darg et al. (2010)
Lintott et al. (2008)
Results

DeepMerge II: Building Robust Deep Learning Algorithms for Merging Galaxy Identification Across Domains

arxiv:2103.01373

Accuracy ~ 50%

<table>
<thead>
<tr>
<th>Source Domain</th>
<th>Target Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>noDA</td>
<td>91%</td>
</tr>
<tr>
<td>MMD</td>
<td>94%</td>
</tr>
<tr>
<td>MMD+T</td>
<td>83%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source Domain</th>
<th>Target Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>noDA</td>
<td>50%</td>
</tr>
<tr>
<td>MMD</td>
<td>53%</td>
</tr>
<tr>
<td>MMD+T</td>
<td>69%</td>
</tr>
</tbody>
</table>
Results

MMD with TL form previous model - works!

Accuracy ~ 70%

<table>
<thead>
<tr>
<th>Source Domain</th>
<th>Target Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>noDA</td>
<td>91%</td>
</tr>
<tr>
<td>MMD</td>
<td>94%</td>
</tr>
<tr>
<td>MMD+T</td>
<td>83%</td>
</tr>
</tbody>
</table>
Summary

- **Merging galaxies** are important for the study of galaxy morphology, but also evolution of structure in the Universe.
- **Domain adaptation (DA) is crucial** for successful bridging between different data sets and full utilisation of ML in science.

What’s next?

- Harder problems will need more sophisticated methods that try to align classes (MMD aligns the entire distribution).
- Discrepant domains can lead to negative transfer and impact the performance.
- Can DA help us make more robust algorithms, understand decision boundaries and uncertainties of our ML algorithms?
Thank you!

aleksand@fnal.gov