

# Domain Adaptation By Feature Augmentation

Murhaf Hossari

Based on publication: Frustratingly Easy Domain  
Adaptation

By Hal Daumé III

# Outline

- Introduction
- Problem
- Previous Work
- Approach
- Main Findings

# Introduction: Machine Learning

Supervised Learning

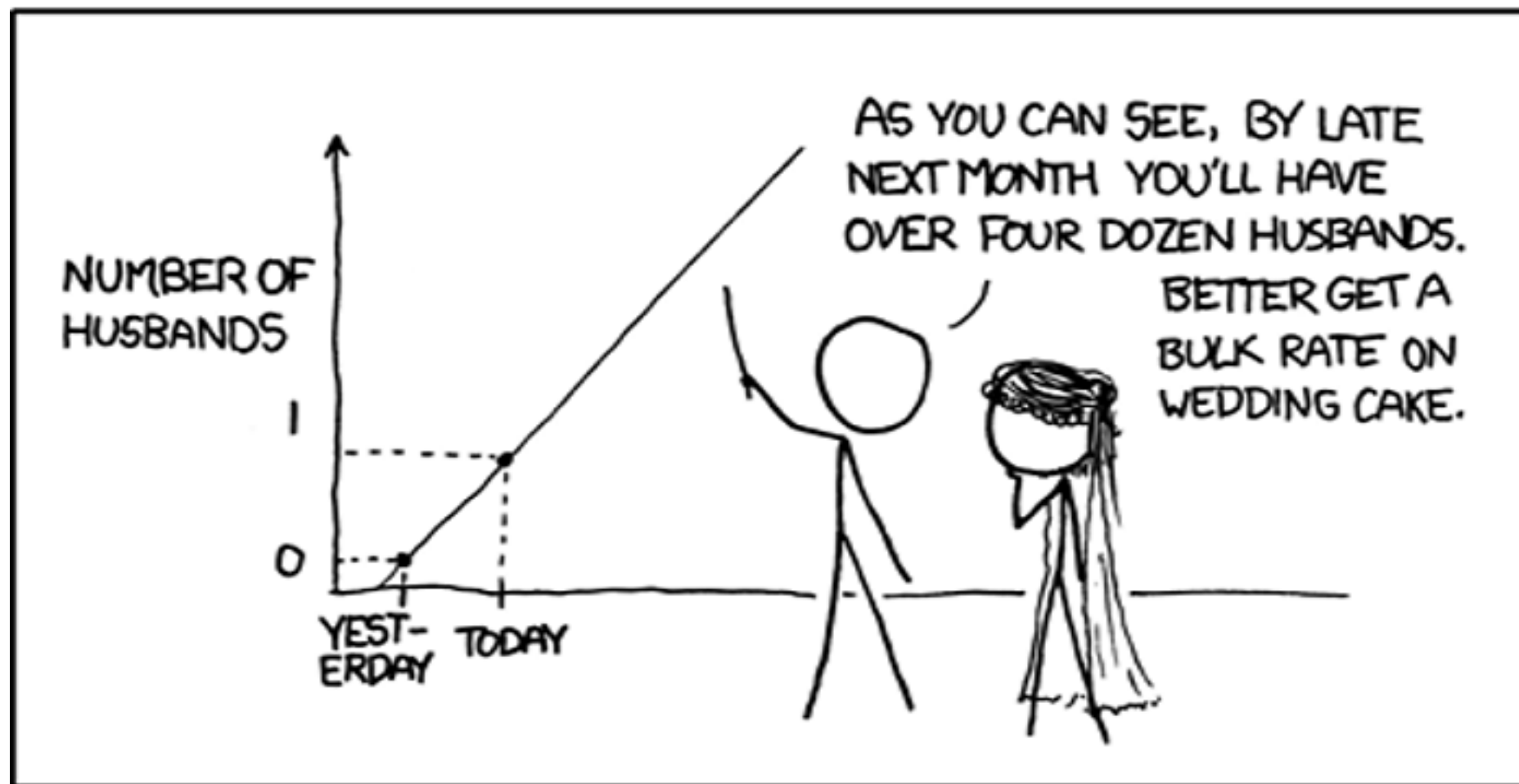


Unsupervised Learning

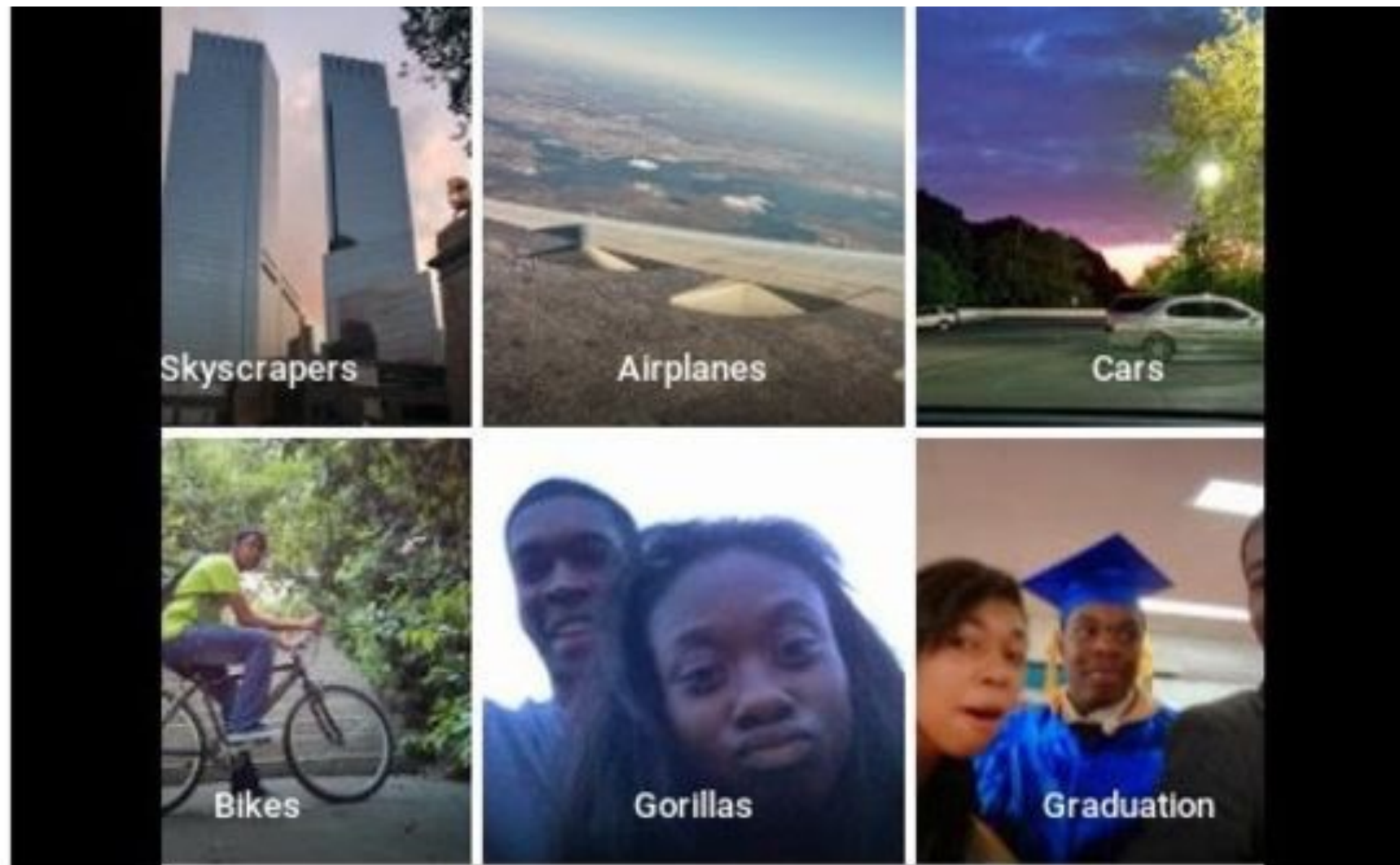


# Introduction: Machine Learning - Supervised

MY HOBBY: EXTRAPOLATING



# Problem: ML goes wrong



**diri noir avec banan** @jackyalcine · Jun 29

Google Photos, y'all [REDACTED] My friend's not a gorilla.



813



394



TWITTER

# Problem: ML goes wrong



**paultag** @paultag · Jun 29

.@googlephotos Y'all need to apologize and own this issue.



**Roan the Goan** @RoanTheGoan · 6 hrs

Oh boy!! is that the first instance of A.I **racism** #fauxpas #GoogleGoneWrong  
#Google



**Tom Sciortino** @SciortinoTom · Jun 30

I'd say that program isn't **racist** so much as it's **stupid**. Granted, there's not much difference between the two.

TWITTER



**diri noir avec banan**

@jackyalcine



Follow

What kind of sample image data you collected that would result in this son?

RETWEETS  
25

FAVORITES  
28



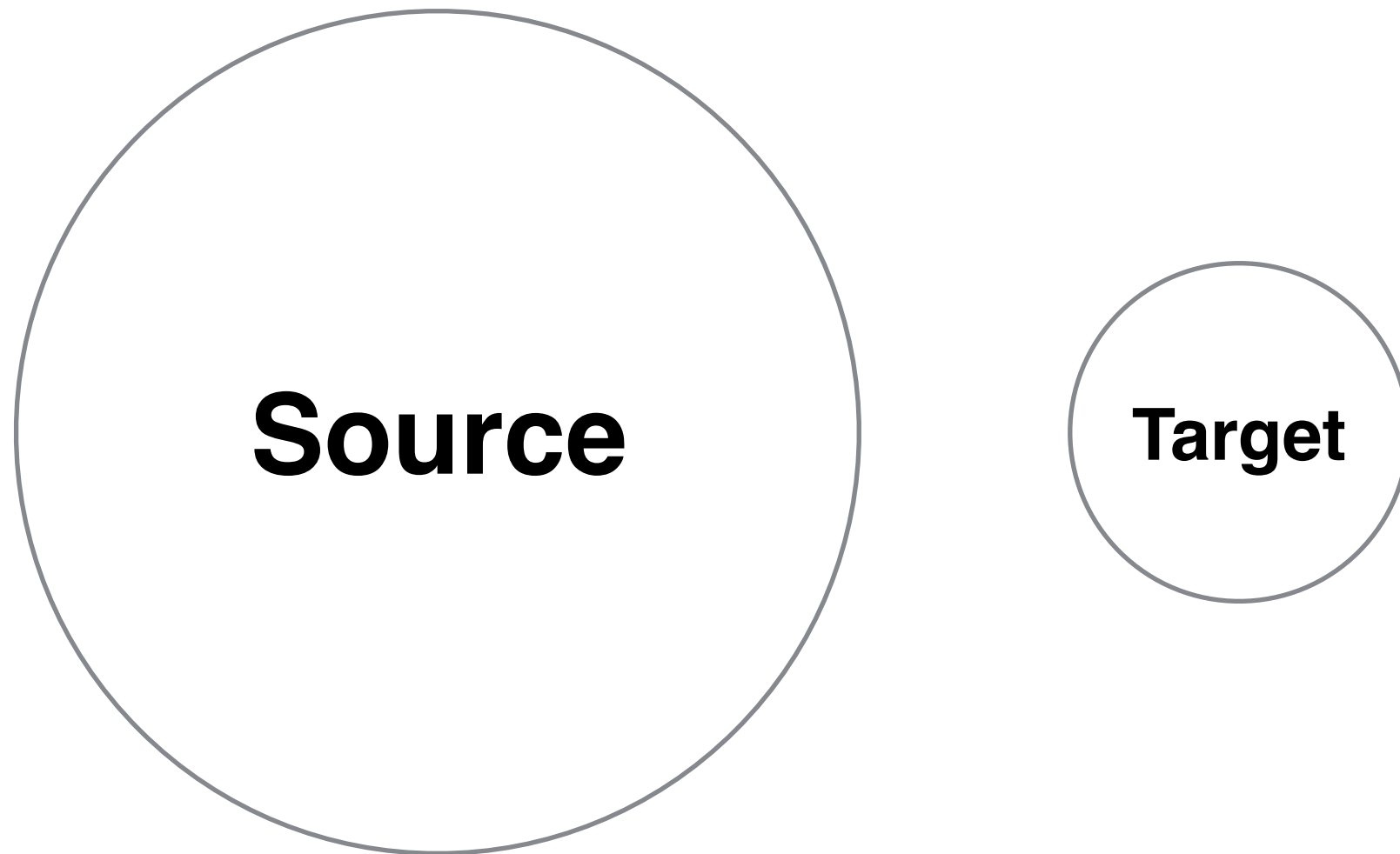
6:26 PM - 28 Jun 2015



# Problem: Domain Difference

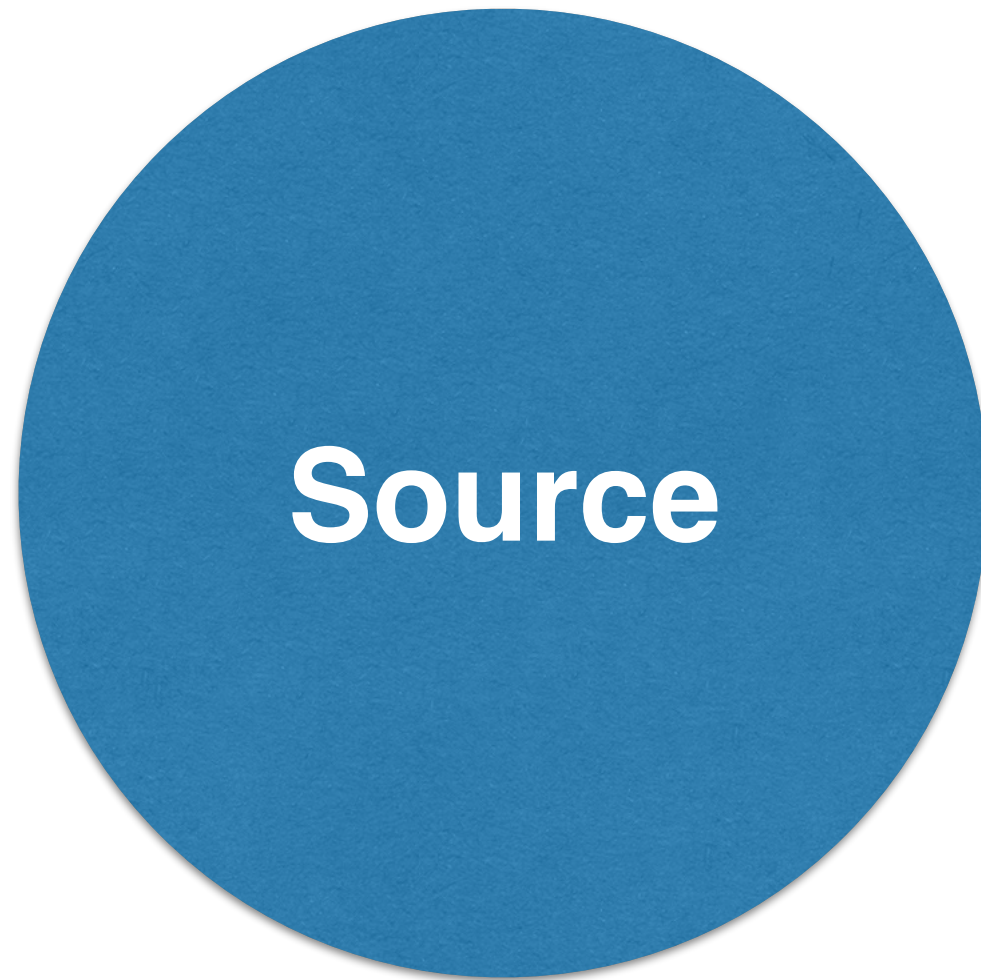


# Litrature: Baselines

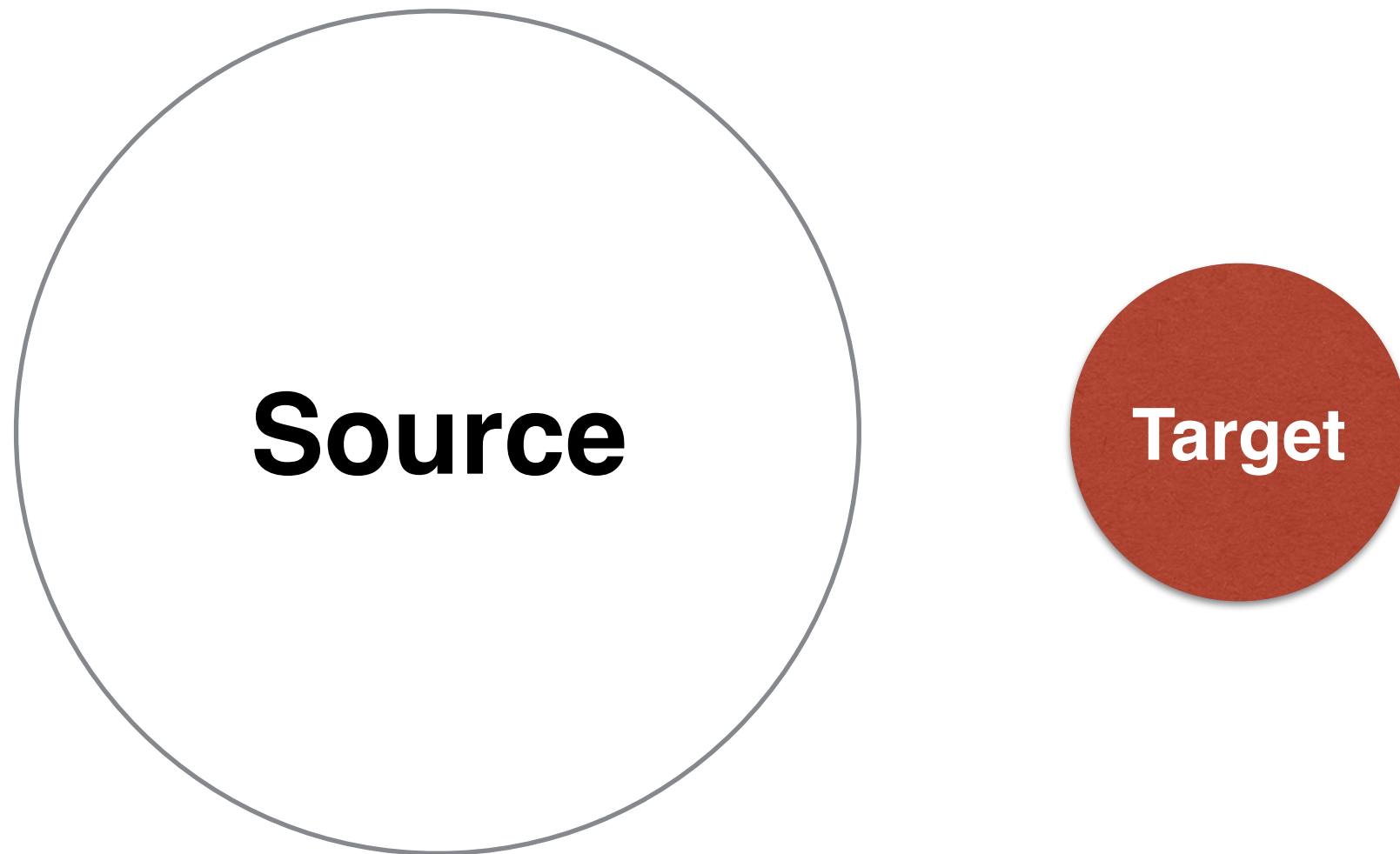




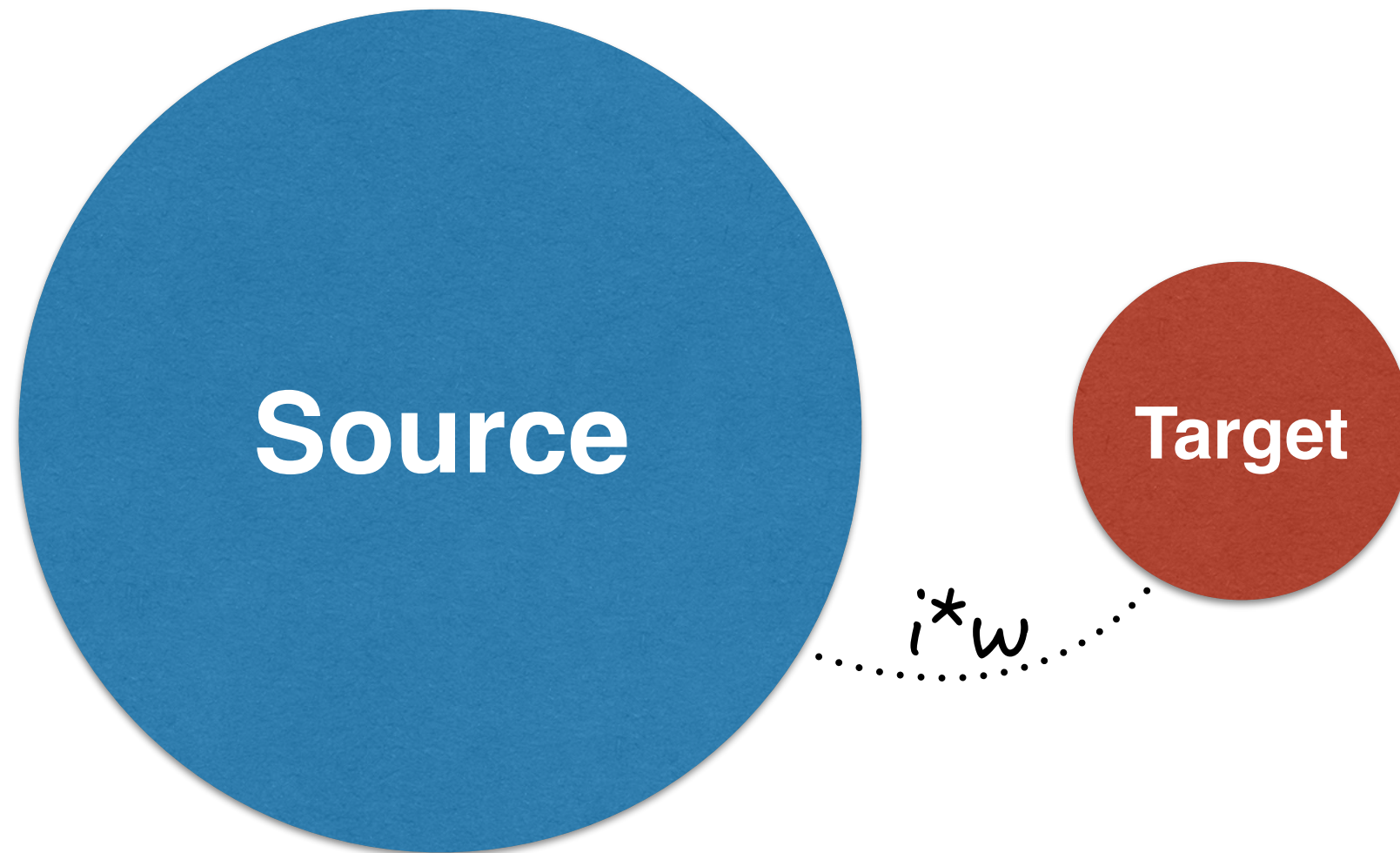
# Litrature: Baselines



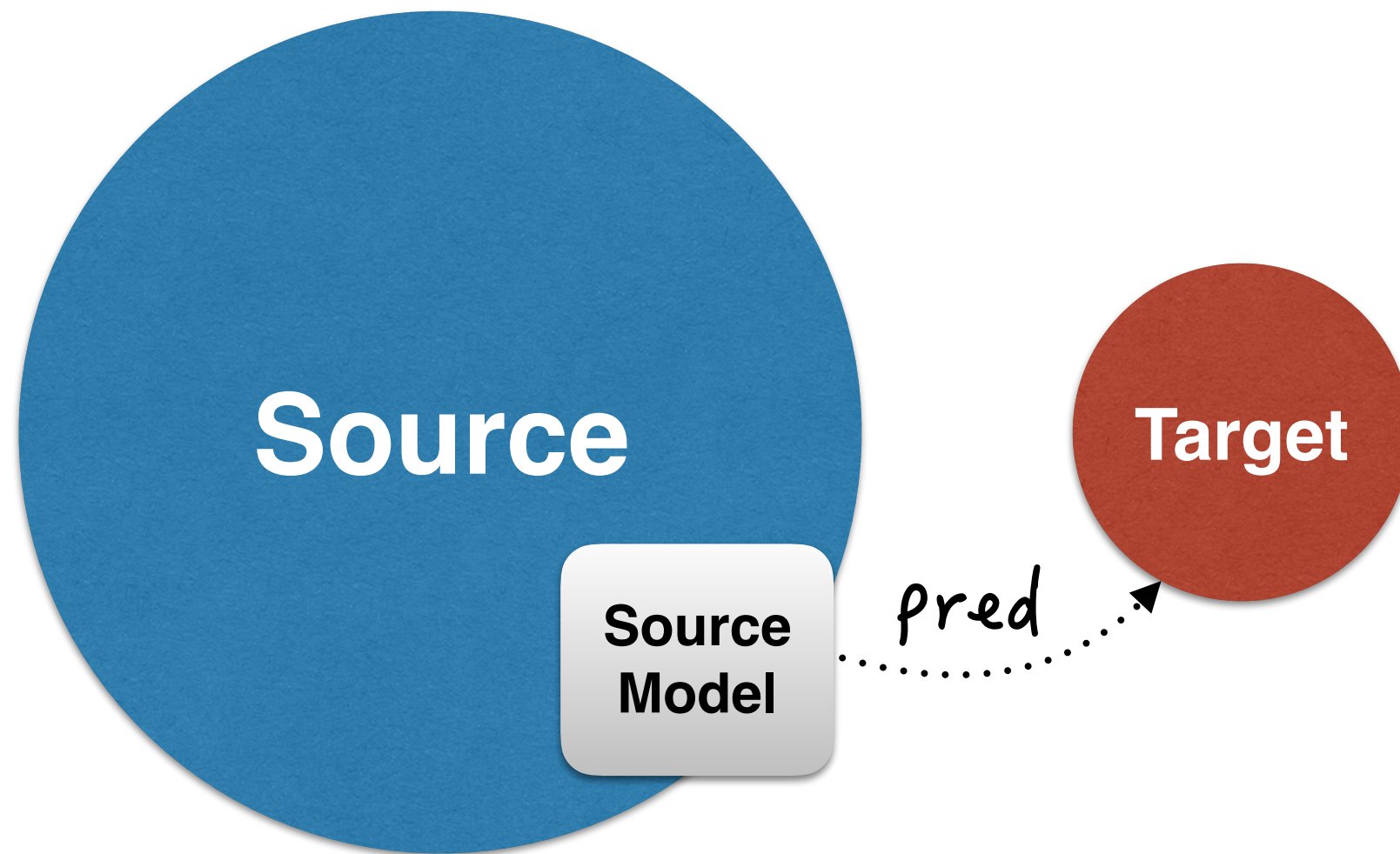
# Litrature: Baselines



# Litrature: Baselines

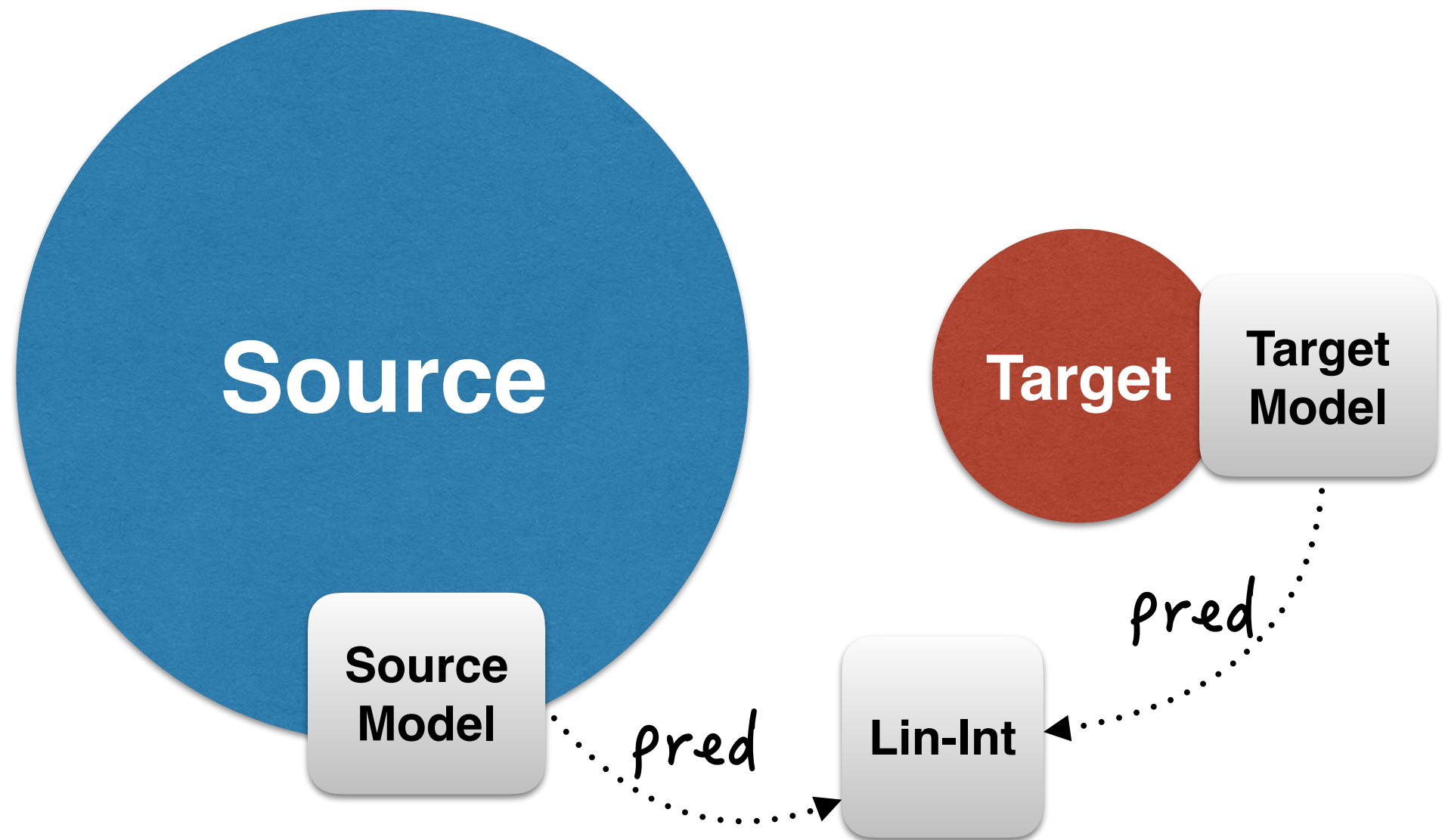


# Litrature: Baselines

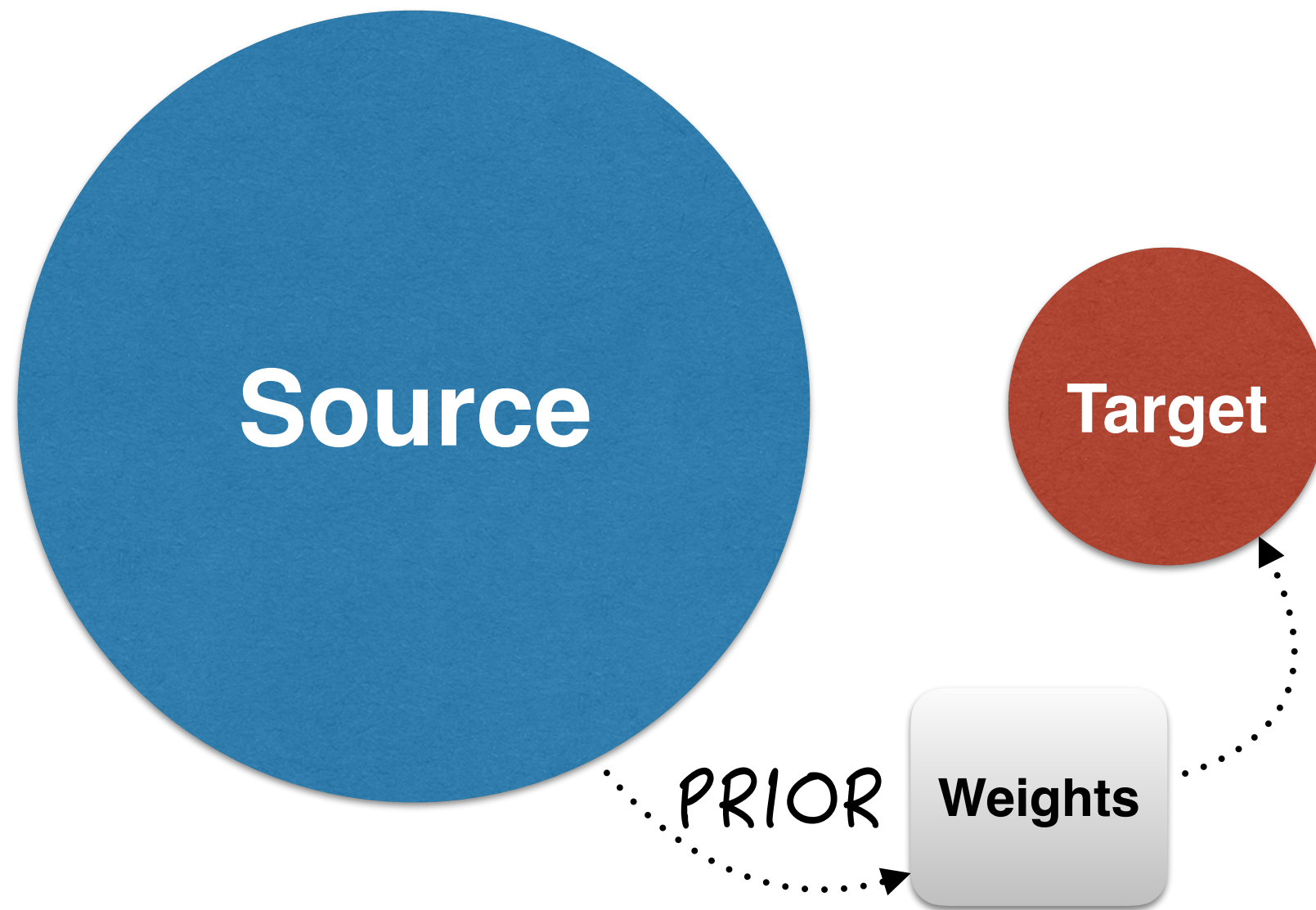




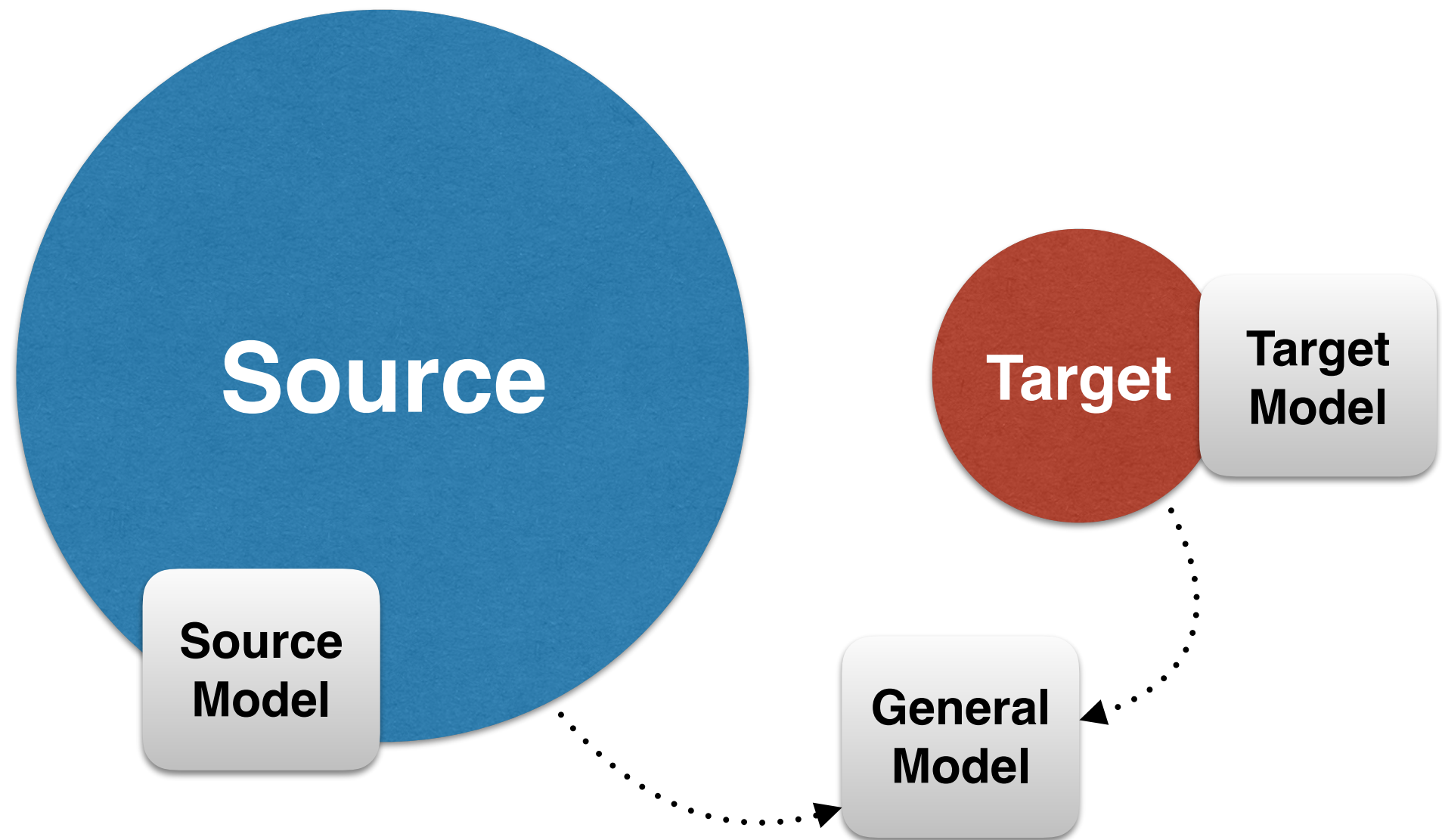
# Litrature: Baselines



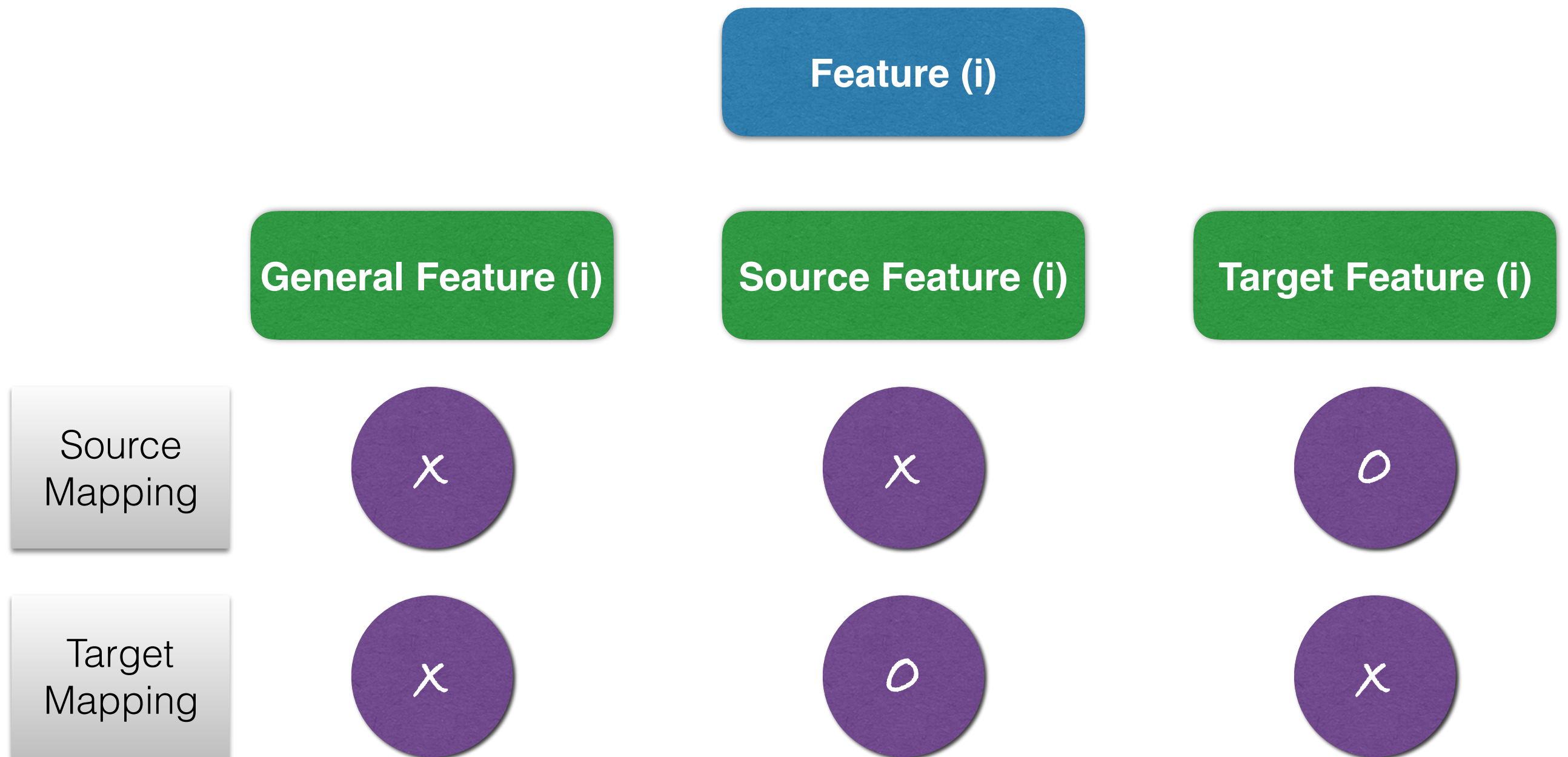
# Literature: Better approaches



# Literature: Better approaches

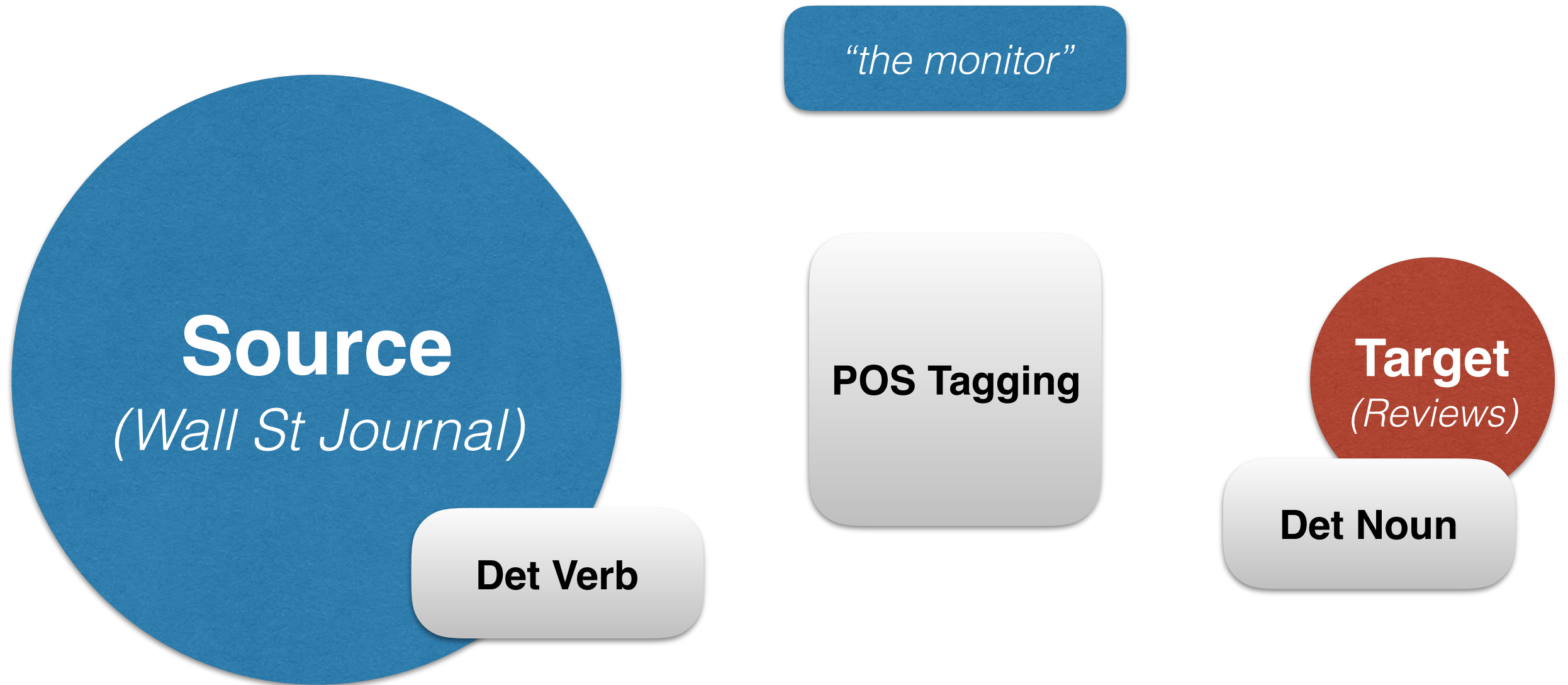


# Approach: Feature Augmentation





# Approach: Example



# Approach: Example

	$(x1, x2)$	the monitor	
	General	Source	Target
	$x1, x2$	$x3, x4$	$x5, x6$
Det (Weights)	1, 0	0, 0	0, 0
Noun (Weights)	0, 0	0, 0	0, 1
Verb (Weights)	0, 0	0, 1	0, 0

# Example Usage

*Word Segmentation of Informal Arabic with  
Domain Adaptation*

*By **Will Monroe, Spence Green, and Christopher D.  
Manning (2014)***

# Main Findings

- Makes learning easier. Less complexity.
- Similar to PRIOR with advantages.
- Experiments ran against sequence labelling tasks.
- Datasets from different domains.
- **Results:** out performs baselines and PRIOR (not in cases where domains are very similar).



*Questions?*

–Murhaf Hossari