University of BRISTOL

9.4 Transfer

Edwin Simpson

Department of Computer Science,

University of Bristol, UK.

bristol.ac.uk

Transfer

- Pretraining a model on one task and then applying it to another:

- Frozen embeddings are **direct transfer** as no adaptation is made.
- Fine-tuning is a kind of **inductive transfer learning** as the model is adapted by further learning on the target task.

Motivations: Human Learning

An Overview of Multi-Task Learning in Deep Neural Networks, blog by Sebastian Ruder, 2017

- When we learn a new task, we apply knowledge of related tasks we have learned before.
- Babies learn to recognise faces first, then apply this to recognise objects.

- When learning to cook a recipe, we transfer our knowledge of how to perform basic tasks like cutting vegetables, frying…

Motivations: Inductive Bias

An Overview of Multi-Task Learning in Deep Neural Networks, blog by Sebastian Ruder, 2017

- Inductive bias: the learner's preference for certain solutions.
- E.g., a NN is more likely to learn certain weights than others

Source task

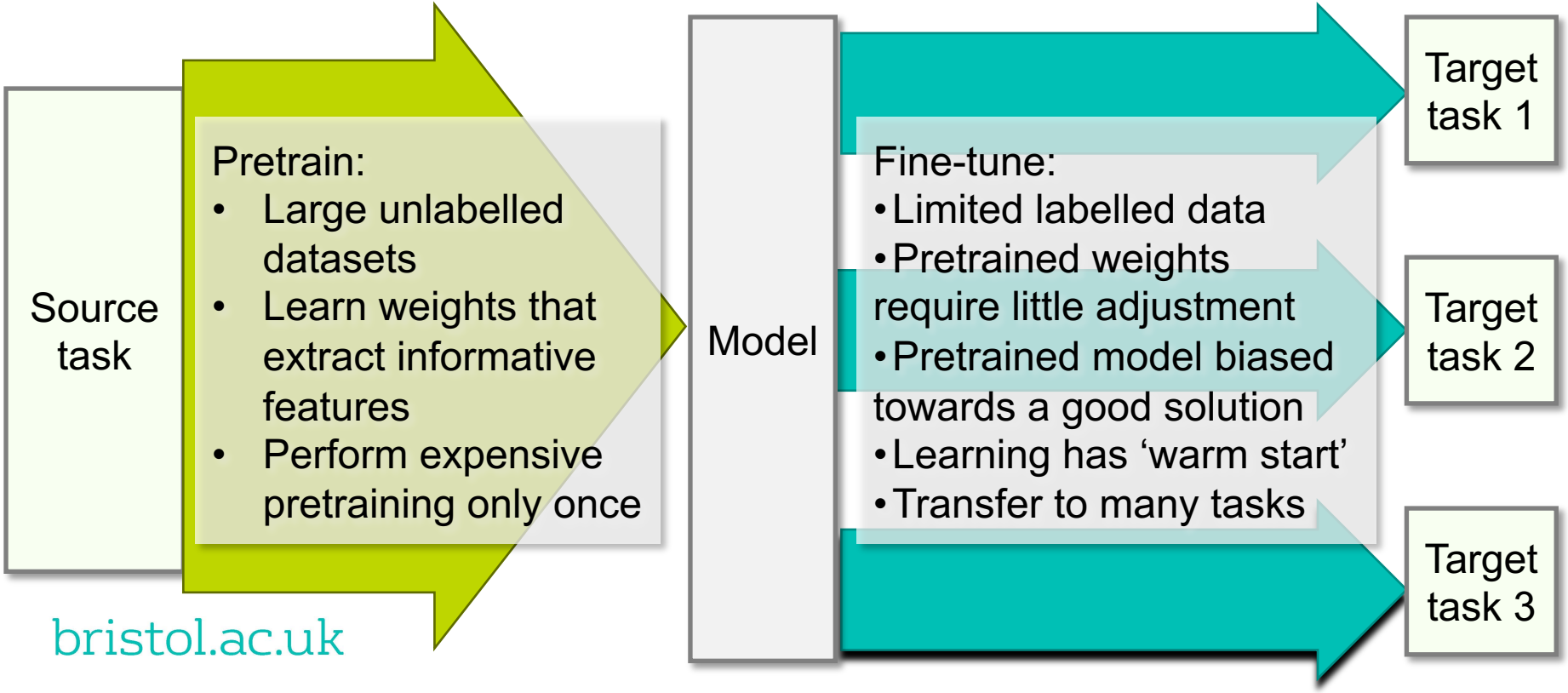
Pretrain: Induce a bias into the model toward good solutions for target task.

Model

Fine-tune: Transfer inductive bias to target task so learner finds a good solution with less training data.

Target task

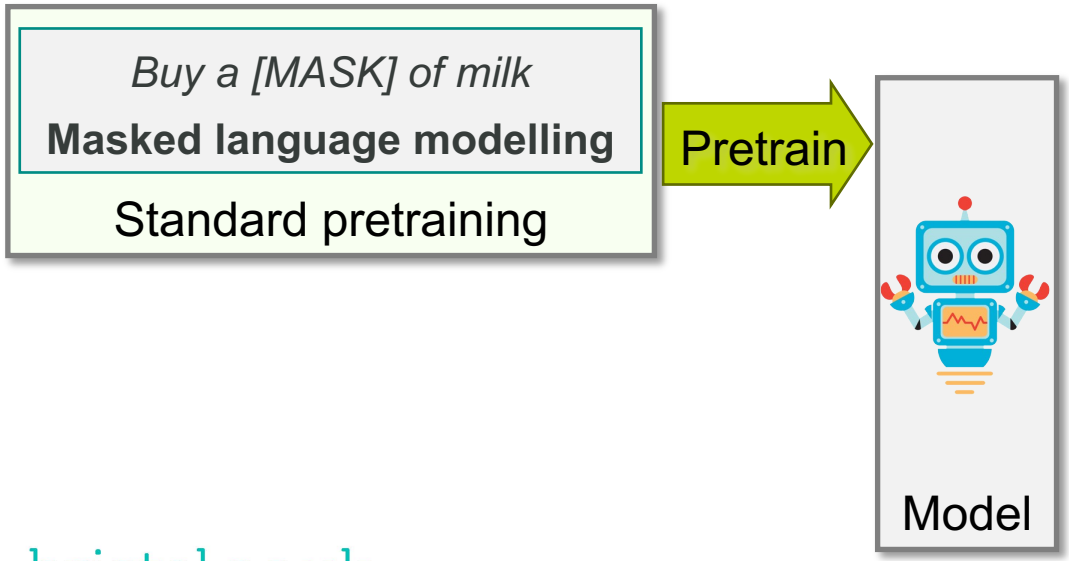
Motivations: Pretrained Embeddings



Domain Adaptation

Don't Stop Pretraining: Adapt Language Models to Domains and Tasks, Gururangan et al., 2020

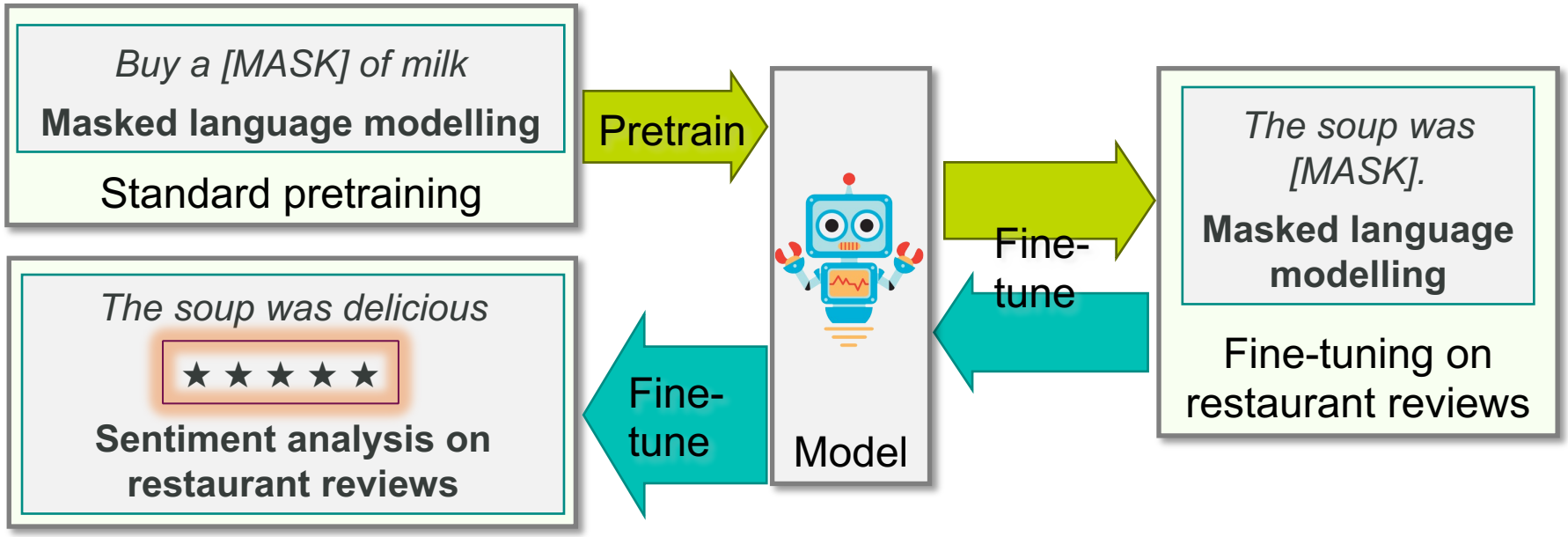
- Fine-tune the language model to the type of text (domain) of your target task



Domain Adaptation

Don't Stop Pretraining: Adapt Language Models to Domains and Tasks, Gururangan et al., 2020

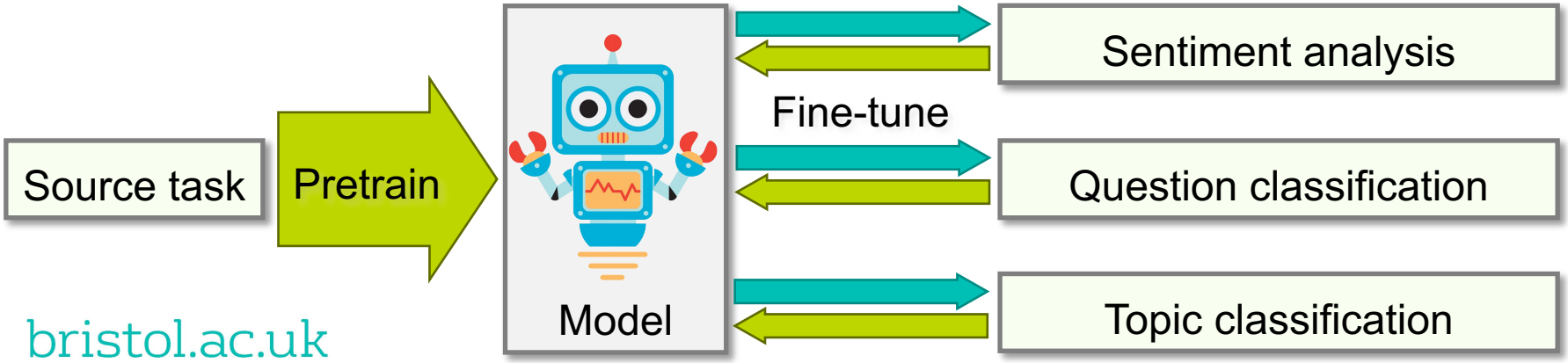
- Fine-tune the language model to the type of text (domain) of your target task



Multi-task Learning

An Overview of Multi-Task Learning in Deep Neural Networks, Ruder, 2017.
How to Fine-Tune BERT for Text Classification? Sun et al., 2020.

- Good features tend to generalise across related tasks
- Training on multiple tasks at the same time learns more general features from more data and can reduce overfitting
- This transfers and combines inductive bias across tasks



Summary

- Pretraining a model then applying or adapting it to another task is a way of transferring knowledge.
- Transfer allows us to:
  - Leverage large unlabelled datasets for pretraining
  - Use the easiest task for learning a particular feature
  - Share information between related tasks.
- Successful transfer learning approaches include:
  - Multiple pretraining tasks
  - Domain adaptation by fine-tuning embeddings
  - Multi-task learning.