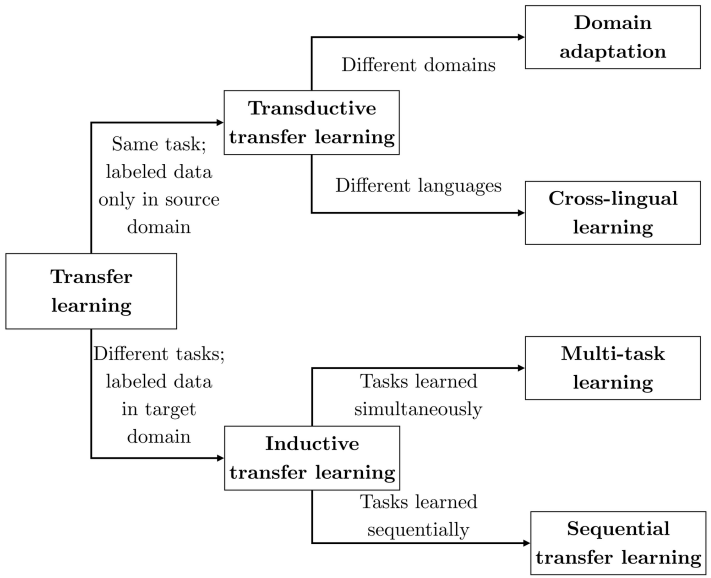
Transfer Learning

Basic definitions and challenges



Learning goals

- Differentiate the different flavors of transfer learning
- Understand the challenges we might be able to overcome by using transfer learning

WHAT IS TRANSFER LEARNING?

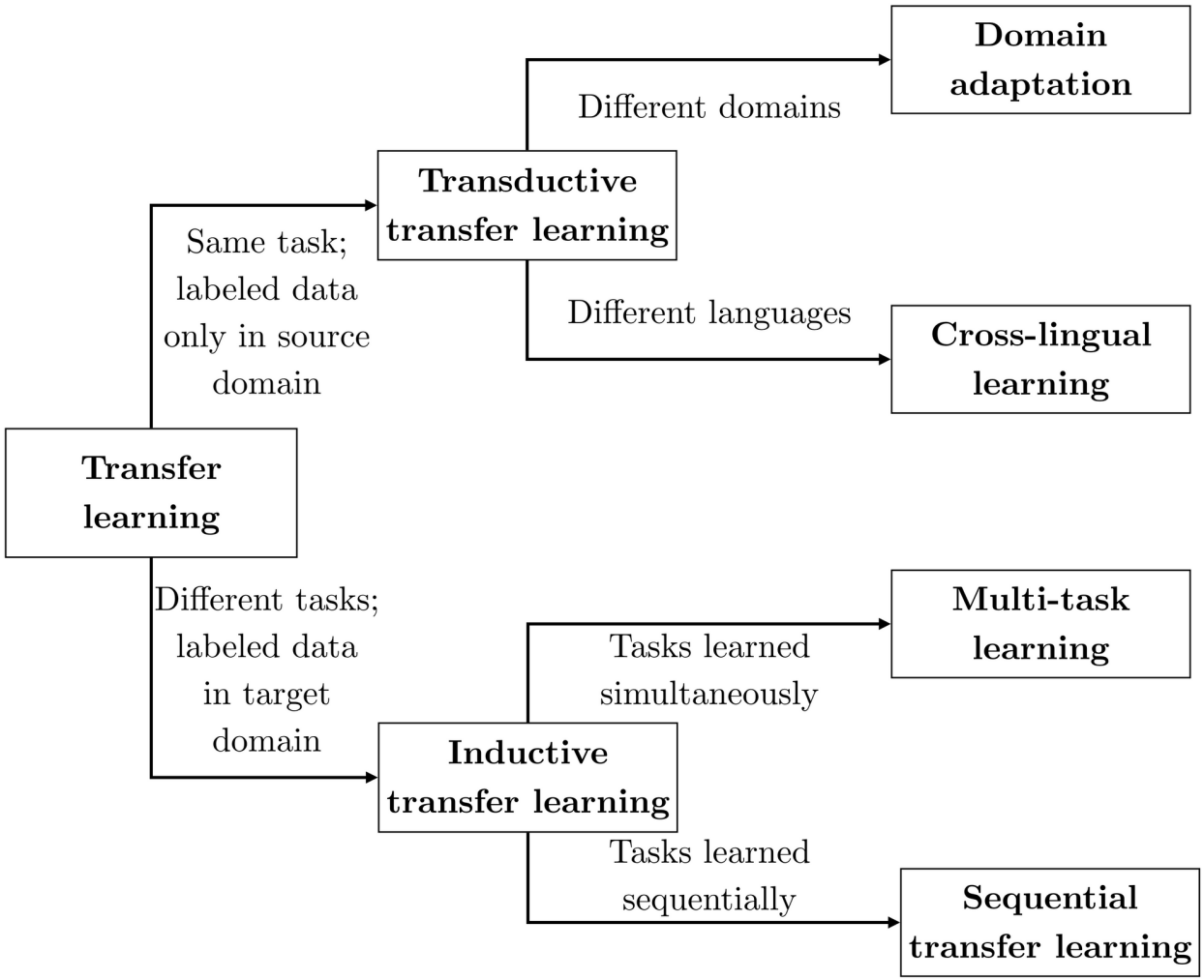
Wikipedia says:
"Transfer learning is a research problem in machine learning that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem."

How it works with word2vec

- Train word2vec on some "fake task" (CBOW or Skip-gram)
- Extract the stored knowledge (a.k.a. embedding)
  or: Directly download embeddings from the web
- Perform a different (supervised) task using the embeddings

TAXONOMY OF TRANSFER LEARNING

▸ RUDER, 2019



Source: *Sebastian Ruder*

TAXONOMY OF TRANSFER LEARNING ▸ RUDER, 2019

Transductive Transfer learning

- Domain adaptation:
  $\rightarrow$ "*Transfer knowledge learned from performing task A on labeled data from domain X to performing task A in domain Y.*"

- Cross-lingual learning:
  $\rightarrow$ "*Transfer knowledge learned from performing task A on labeled data from language X to performing task A in language Y.*"

- *Important:* No labeled data in target domain/language *Y*.

TAXONOMY OF TRANSFER LEARNING ▸ RUDER, 2019

Inductive Transfer learning

- Multi-task learning:
  $\rightarrow$ "*Transfer knowledge learned from performing task A on data from domain X to performing multiple (simultaneous) tasks B, C, D, .. in domain Y.*"

- Sequential transfer learning:
  $\rightarrow$ "*Transfer knowledge learned from performing task A on data from domain X to performing multiple (sequential) tasks B, C, D, .. in domain Y.*"

- *Important:* Labeled data only for task(s) from target domain *Y*.

CHALLENGES I

Low-resource environments:

- The larger the models, the more data is needed to train them
- (Labeled) Data is scarce and expensive!
- Many languages in the world are highly underrepresented in terms of existing resources
  Often: *Number of speakers* $\neq$ *Amount of available written text*
- Unlabeled (English) text data is ubiquitous

CHALLENGES II

Cross-lingual transfer:

- Languages can be grouped into certain families
- Patterns that a model learns for one language, might be beneficial for learning a second language (just as it is for us humans as well: For those who learned French in high school, learning Spanish afterwards might be easier)
- Again: Scarcity of resources; assume the following scenario:
  - **Large** parallel corpus for languages A and B
  - **Large** parallel corpus for languages A and C
  - *Small* parallel corpus for languages B and C

  $\rightarrow$ Training a model for B and C in isolation not the best idea