

– IN5550 –

Neural Methods in Natural Language Processing

Ensembles, transfer and **multi-task learning**: Part 2

Erik Velldal

Language Technology Group (LTG)
University of Oslo



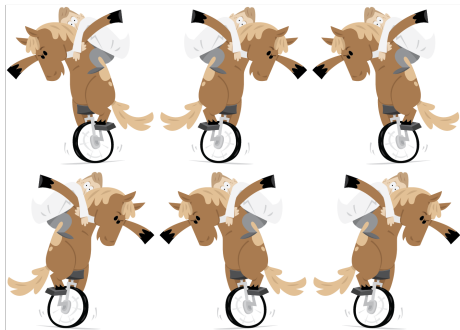
- ▶ Standard single-task models:



- ▶ Standard single-task models:



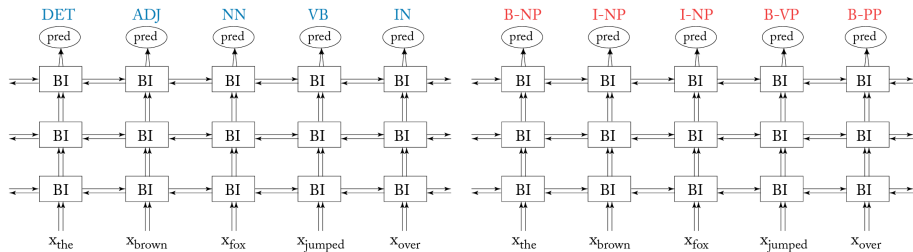
- ▶ Ensembles:



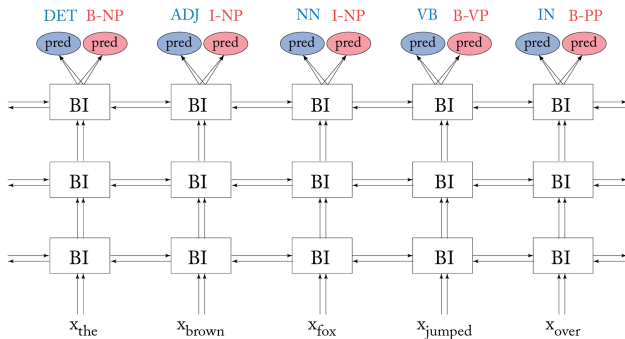


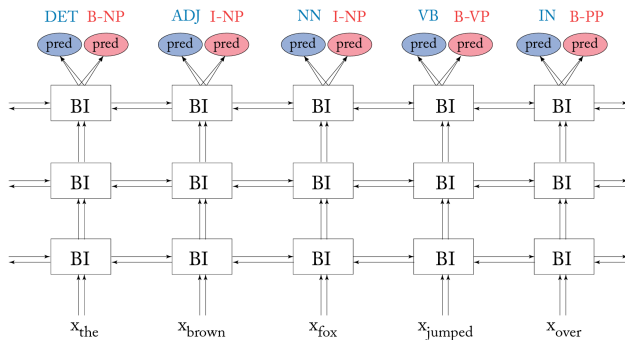
- ▶ Train **one model** to solve **multiple tasks**.
- ▶ Each task has its own loss-function, but the model weights are (partly) shared.
- ▶ Examples for the different labels can be distinct (take turns picking examples) or the same.
- ▶ Most useful for closely related tasks.
- ▶ Example: PoS-tagging and syntactic chunking.

Standard single-model approach



Multi-task approach



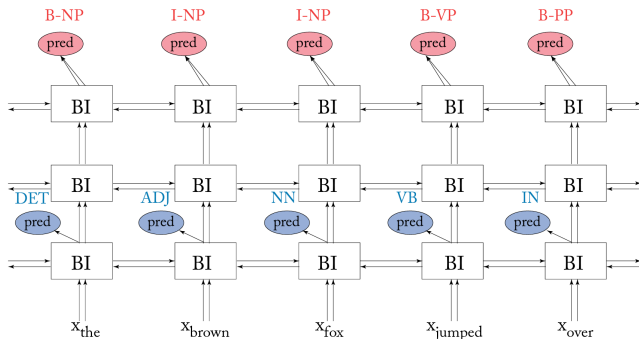


- Often one task will be considered the **main task**; the others so-called supporting- or **auxilliary tasks**.

Hierarchical / cascading multi-task learning

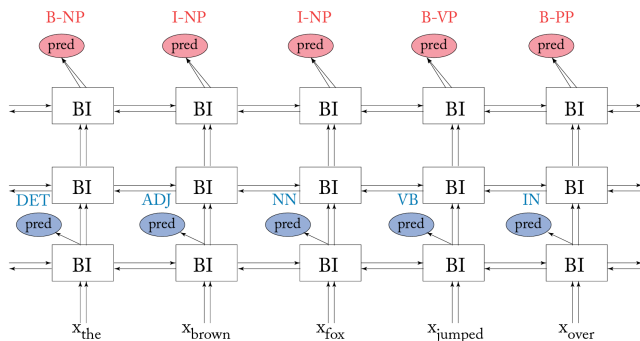


- ▶ Observation: while relying on similar underlying information, tagging intuitively seems more low-level than chunking.
- ▶ Cascading architecture with selective sharing of parameters:





- ▶ Observation: while relying on similar underlying information, tagging intuitively seems more low-level than chunking.
- ▶ Cascading architecture with selective sharing of parameters:



- ▶ Note that the units of classification for the main and aux. tasks can be different, e.g. sentence- vs word-level.



- ▶ Transfer learning