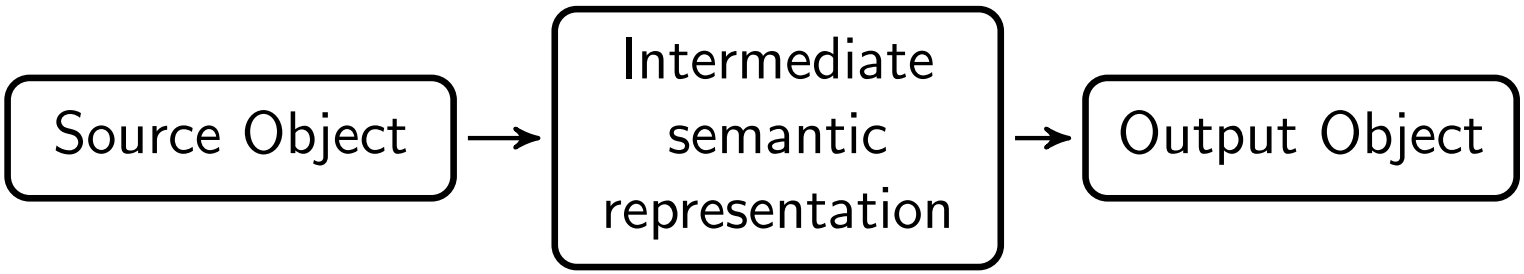
NLP

Machine Translation

St. Joseph's University, Bengaluru

ML 2 BDA3321

Neural Machine Translation



Translation with MLP

The MLP estimates

$$P(t_1, t_2, \ldots, t_n | s_1, s_2, \ldots s_n)$$

Drawbacks

It requires sentences to be preprocessed to be of fixed length.

Structure of the netwerk

1. Model reads the input sequence.
2. Model emits a data structure that summarizes the input sequence. This is called the context C.
3. A second model reads C and generates a sentence in the target language.

Using an attention mechanism

1. A process that **reads** raw data and converts them into distributed representations.
2. A list of feature vectors storing the output of the reader. This can be understood as **memory** containing a sequence of facts.
3. A process that **exploits** the content of the memory to sequentially perform a task, at each step having the ability to put attention on the content of one memory element.

References I

[Mur12] Kevin P Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[Mar14] Stephen Marsland. *Machine Learning, An Algorithmic Perspective*. CRC Press, 2014.

[Ian17] Aaron Courville Ian Goodfellow Yoshua Bengio. *Deep Learning*. MIT Press, 2017.