MacNet: Transferring Knowledge from Machine Comprehension to Sequence-to-sequence Models

Boyuan Pan, Yanzheng Yang, Hao Li, Zhou Zhao, Yueting Zhuang, Deng Cai, Xiaofei He.
Zhejiang University

Introduction
We propose MacNet: a novel encoder-decoder supplementary architecture to the widely used attention-based sequence-to-sequence models, which transfers knowledge learned from machine comprehension to sequence-to-sequence tasks to deepen the understanding of the text.

Machine Comprehension (MC):
- It requires to provide an answer (usually a subspan in the passage) given a passage and a question.
- An example of a typical question-
  answer scenario:

Machine Comprehension Encoding Layer

Attention Layer

Modeling Layer

Output Layer

Encoder: Original encoder + transferred encoder.

Decoder & Attention Mechanism: Additionally send the attention vector into the modeling layer of the pre-trained MC model.

MacNet Architecture

Sequence-to-sequence Model:

Introduction
We propose MacNet: a novel encoder-decoder supplementary architecture to the widely used attention-based sequence-to-sequence models, which transfers knowledge learned from machine comprehension to sequence-to-sequence tasks to deepen the understanding of the text.

Machine Comprehension (MC):
- It requires to provide an answer (usually a subspan in the passage) given a passage and a question.
- An example of a typical question-
  answer scenario:

Machine Comprehension Encoding Layer

Attention Layer

Modeling Layer

Output Layer

Encoder: Original encoder + transferred encoder.

Decoder & Attention Mechanism: Additionally send the attention vector into the modeling layer of the pre-trained MC model.

MacNet Architecture

Sequence-to-sequence Model:

Introduction
We propose MacNet: a novel encoder-decoder supplementary architecture to the widely used attention-based sequence-to-sequence models, which transfers knowledge learned from machine comprehension to sequence-to-sequence tasks to deepen the understanding of the text.

Machine Comprehension (MC):
- It requires to provide an answer (usually a subspan in the passage) given a passage and a question.
- An example of a typical question-
  answer scenario:

Machine Comprehension Encoding Layer

Attention Layer

Modeling Layer

Output Layer

Encoder: Original encoder + transferred encoder.

Decoder & Attention Mechanism: Additionally send the attention vector into the modeling layer of the pre-trained MC model.

MacNet Architecture

Sequence-to-sequence Model:

Introduction
We propose MacNet: a novel encoder-decoder supplementary architecture to the widely used attention-based sequence-to-sequence models, which transfers knowledge learned from machine comprehension to sequence-to-sequence tasks to deepen the understanding of the text.

Machine Comprehension (MC):
- It requires to provide an answer (usually a subspan in the passage) given a passage and a question.
- An example of a typical question-
  answer scenario:

Machine Comprehension Encoding Layer

Attention Layer

Modeling Layer

Output Layer

Encoder: Original encoder + transferred encoder.

Decoder & Attention Mechanism: Additionally send the attention vector into the modeling layer of the pre-trained MC model.

MacNet Architecture

Sequence-to-sequence Model:

Introduction
We propose MacNet: a novel encoder-decoder supplementary architecture to the widely used attention-based sequence-to-sequence models, which transfers knowledge learned from machine comprehension to sequence-to-sequence tasks to deepen the understanding of the text.

Machine Comprehension (MC):
- It requires to provide an answer (usually a subspan in the passage) given a passage and a question.
- An example of a typical question-
  answer scenario:

Machine Comprehension Encoding Layer

Attention Layer

Modeling Layer

Output Layer

Encoder: Original encoder + transferred encoder.

Decoder & Attention Mechanism: Additionally send the attention vector into the modeling layer of the pre-trained MC model.

MacNet Architecture

Sequence-to-sequence Model:

Conclusion
We improve the sequence-to-sequence model via transferring knowledge of several neural network layers from another supervised task.
- We conduct extensive experiments on two typical seq2seq tasks to show that our method achieves significant improvement on the baseline model.