

## 论文摘要

近年来, 神经机器翻译模型已经成为机器翻译领域的主流模型, 如何从大量的训练数据中快速、准确地学习翻译知识是一个值得探讨的问题。不同的训练样本的难易程度不同, 有的训练样本比较简单, 易于模型学习, 而有的训练样本比较难, 不易于模型学习, 样本的难易程度对模型的收敛性有极大影响, 但是传统的神经机器翻译模型在训练过程中并没有考虑这种差异性。因此, 本文探究了样本的难易程度对神经机器翻译模型的训练过程的影响, 具体的, 我们基于“课程学习”的思想, 为神经机器翻译模型提出了一种基于样本难度的动态学习方法。我们分别从神经机器翻译模型的翻译效果和训练样本的句子长度两个方面来量化训练样本的难易程度, 然后设计了由易到难的和由难到易的两种学习策略来训练模型, 并比较模型的翻译效果。实验结果表明, 由易到难和由难到易的动态学习方法均可以提高神经机器翻译模型的翻译效果。

## 系统模型

模型的整体架构是基于注意力机制的编码器-解码器结构, 主要包括三个部分: 编码器、解码器和注意力网络。

## 论文简介

本文针对神经机器翻译不同训练数据之间难易程度的差异性, 借鉴“课程学习”的思想, 为神经机器翻译模型提出了一种基于样本难度的动态学习方法。

## 算法原理

用  $D_{(x,y)}$  表示训练样本  $(x,y) \in D$  的难易程度,  $D_{(x,y)} \in [0,1]$ 。  $D_{(x,y)}$  越大, 表示训练样本  $(x,y)$  越简单。用  $W_{(x,y)}^i$  表示训练样本  $(x,y)$  在第  $i$  轮 ( $i$  从 0 开始计数) 训练过程中的权重, 该权重基于训练样本的难易程度和训练轮数动态变化, 以此来调整模型的训练过程。模型在训练过程中的优化目标变为最小化损失函数:

$$J = \sum_{s=1}^S -(1 + W_{(x,y)}) \log P(y^{(s)} | x^{(s)}; \theta)$$

## 实验仿真

本文分别基于训练样本的 BLEU 值、源语言句子的长度、目标语言句子的长度以及源语言句子和目标语言句子的总长度来量化训练样本的难易程度, 将选用这些样本难度准则的模型依次记作  $M_{BLEU}$ 、 $M_{src}$ 、 $M_{tgt}$  和  $M_{sum}$ , 同时将标准的忽略训练样本难易程度的模型记作  $M_{base}$ 。分别用由易到难的和由难到易的学习策略来调整模型的训练过程, 计算模型在测试集上的 BLEU 值, 实验结果如表 1-1 和 1-2 所示。

表 1-1 由易到难学习策略下的 BLEU 值

模型	BLEU	$\Delta$
$M_{base}$	20.85	-
$M_{BLEU}$	25.08	4.23
$M_{src}$	26.10	5.25
$M_{tgt}$	<b>26.36</b>	<b>5.51</b>
$M_{sum}$	26.30	5.45

表 1-2 由难到易学习策略下的 BLEU 值

模型	BLEU	$\Delta$
$M_{base}$	20.85	-
$M_{BLEU}$	25.77	4.92
$M_{src}$	<b>26.13</b>	<b>5.28</b>
$M_{tgt}$	26.02	5.17
$M_{sum}$	25.85	5.00

## 论文结论

由表 1-1 和 1-2 中的结果可知, 和神经机器翻译模型标准的忽略样本难易程度的学习方法相比, 基于训练样本的难易程度, 分别采取由易到难和由难到易的动态学习策略, 均可以提高模型在测试集上的 BLEU 值, 说明本文提出的基于样本难度的动态学习方法确实可以提高模型的翻译效果。另外, 不同的样本难度准则、不同的动态学习策略对模型翻译效果的提升程度不同, 因此选取合适的样本难度准则和动态学习策略很关键。