



西安电子科技大学
XIDIAN UNIVERSITY

基于多任务学习的文本表示方法研究

Language Representation with Multi-Task Learning

孙天祥

txsun@stu.xidian.edu.cn

2019年4月

Outline



- 毕业设计完成情况
- 课题简介与实验结果
- 存在问题与解决方案
- 下一步计划



一、毕业设计完成情况

- **调研** 学习已有工作、翻译外语文献
- **设计** 设计新的多任务学习模型
- **实现** 实现基线模型及新提出的模型
- **实验** 搭建实验环境、对比试验
- **改进** 根据存在的问题和实验结果改进模型
- **写作** 完成毕业论文

二、课题简介



➤ 题目 基于多任务学习的文本表示方法研究

Language Representation with Multi-Task Learning

1. Multi-Task Learning
2. Natural Language Learning
3. MTL for NLP
4. Transformer

Multi-Task Learning

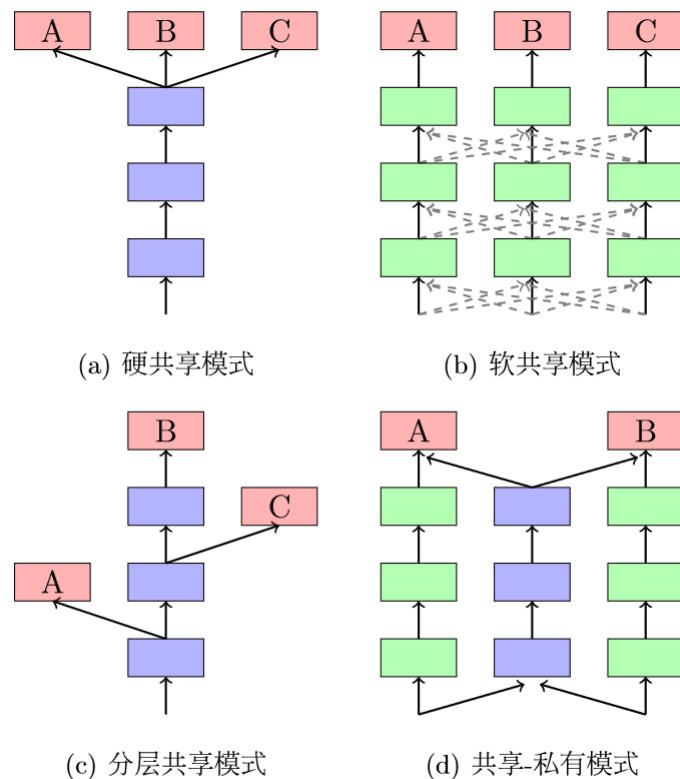


Figure 1.2: 多任务学习的几种常见模式

多任务学习通过利用包含在相关任务训练信号中的领域特定信息来提升泛化能力^[1]。

[1] R. Caruana, "Multitask Learning," Machine Learning, vol. 28, no. 1, pp. 41–75, 1997.

Multi-Task Learning



深层神经网络多任务学习综述（中文翻译）

Sebastian Ruder†

Insight Centre for Data Analytics, NUI Galway‡

孙天祥 ‡

Xidian University‡

Abstract

多任务学习在自然语言处理、语音识别、计算机视觉和药物设计等多个机器学习应用上取得了巨大成功。本文旨在对多任务学习，特别是深层神经网络中的多任务学习，进行一般性的综述。本文介绍了深度学习中的多任务学习的两种最常见的框架，概括和讨论了相关文献和最新进展。最终，本文希望能够通过阐释多任务学习的工作机理来帮助机器学习研究人员在实践中应用多任务学习，并为选择合适的辅助任务提供了指导性建议。

1. 引言

在机器学习中，我们一般关心如何优化一个特定指标，这个指标可能是某个基准测试的分数，也可能是商业 KPI。为此，我们一般训练单个模型或者一组集成模型来完成这个任务。然后，我们会微调这些模型知道它们的性能不再提升。虽然我们通常可以通过这种方式获得可接受的效果，但是由于局限于我们关注的

计算机视觉 [19] 和药物设计 [34]。多任务学习有很多形式：联合学习、学习如何学习、通过辅助任务学习等，这些概念只不过是多任务学习的不同名称。一般地，只要你在优化多个损失函数，你就已经在进行多任务学习了。在这些场景中，显式地考虑多任务学习的概念和机理对于提升性能会有所帮助。

即使你只是像通常设定的那样在优化一个损失函数，也有可能存在能够帮助提升主任务性能的辅助任务。Caruana 等人 [11] 对多任务学习给出了一个简洁的总结：“多任务学习通过利用包含在相关任务训练信号中的领域特定信息来提升泛化能力”。

在本文中，我们将介绍多任务学习的研究现状，特别是当其应用在深度神经网络中的情况。首先，我们将在第 2 节中从不同的视角来解释多任务学习，然后我们将在第 3 节中介绍深度学习中的两种最常见的多任务学习方法，在第 4 节，我们将描述使得多任务学习在实践中奏效的机制。在介绍最近的基于神经网络的多任务学习方法之前，我们会先在第 5 节讨论非神经网络的多任务学习。在第 6 节中将会介绍近期提出的强大的基于神经网络的多任务学习方法。最后，我们将在第 7 节中介绍被用到的辅助任务并讨论如何构造

6.1. 深层关系网络

在计算机视觉的多任务学习方法中，通常共享卷积层，同时学习特定于任务的全连接层。Long and Wang 等人 [29] 通过提出深层关系网络来改进这些模型。除了共享和任务特定层的结构（如图 3 所示），它们将矩阵先验放置在全连接层上，这允许模型学习任务之间的关系。然而，这种方法仍然依赖于预先定义的共享结构，虽然可能足以解决论文中研究的计算机视觉问题，但对于新任务而言非常容易出错。

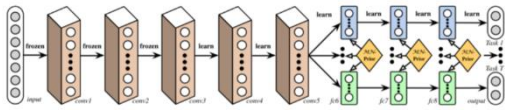


图 3. 深层关系网络

6.2. 完全自适应的特征共享

在另一方面，Lu 等人 [31] 提出了一种自下而上的方法，该方法从瘦网络开始，在训练期间使用促进类似任务分组的标准贪心地扩展此网络。可以在图 4 中看到动态创建分支的网络扩展过程。但是，贪心的方法可能无法找到全局最优的模型，此外将每个分支分配给一个特定的任务也导致模型无法学习任务之间更复杂

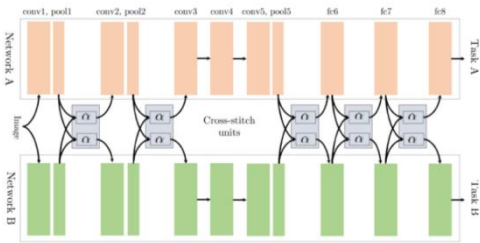


图 5. 十字绣网络

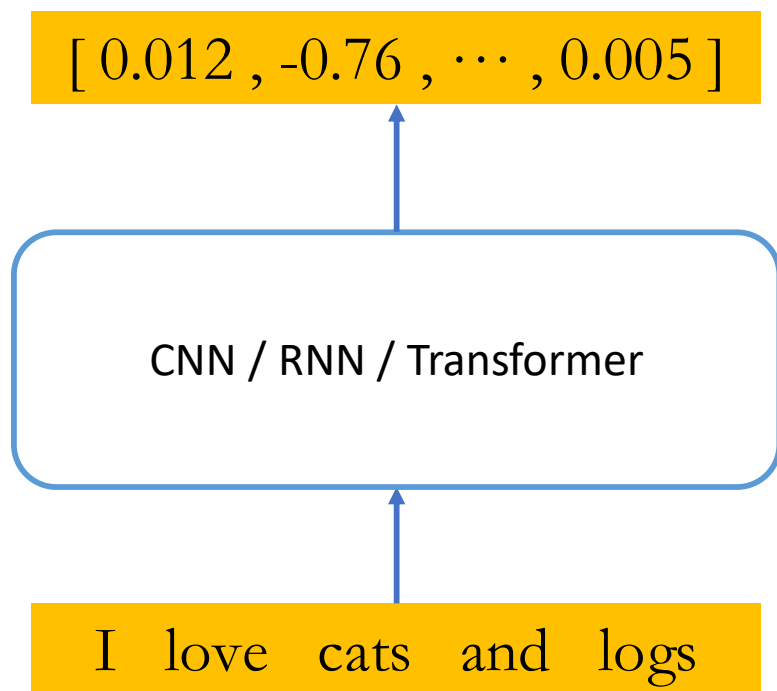
6.4. 分层监督

相比之下，在自然语言处理（NLP）中，近期的工作更多地关注如何为多任务学习找到更好的任务结构层次：Søgaard 和 Goldberg 等人 [38] 通过实验表明当低级别的 NLP 任务（即通常用于预处理的 NLP 任务，如词性标注和命名实体识别）用作辅助任务时，应在较低层进行监督。将监督信号分层施加的模式也被称为分层共享。

6.5. 联合多任务学习模型

基于上面的发现，Hashimoto 等人 [20] 为多个 NLP 任务预先定义了层次结构，如图 6 所示。

Natural Language Processing



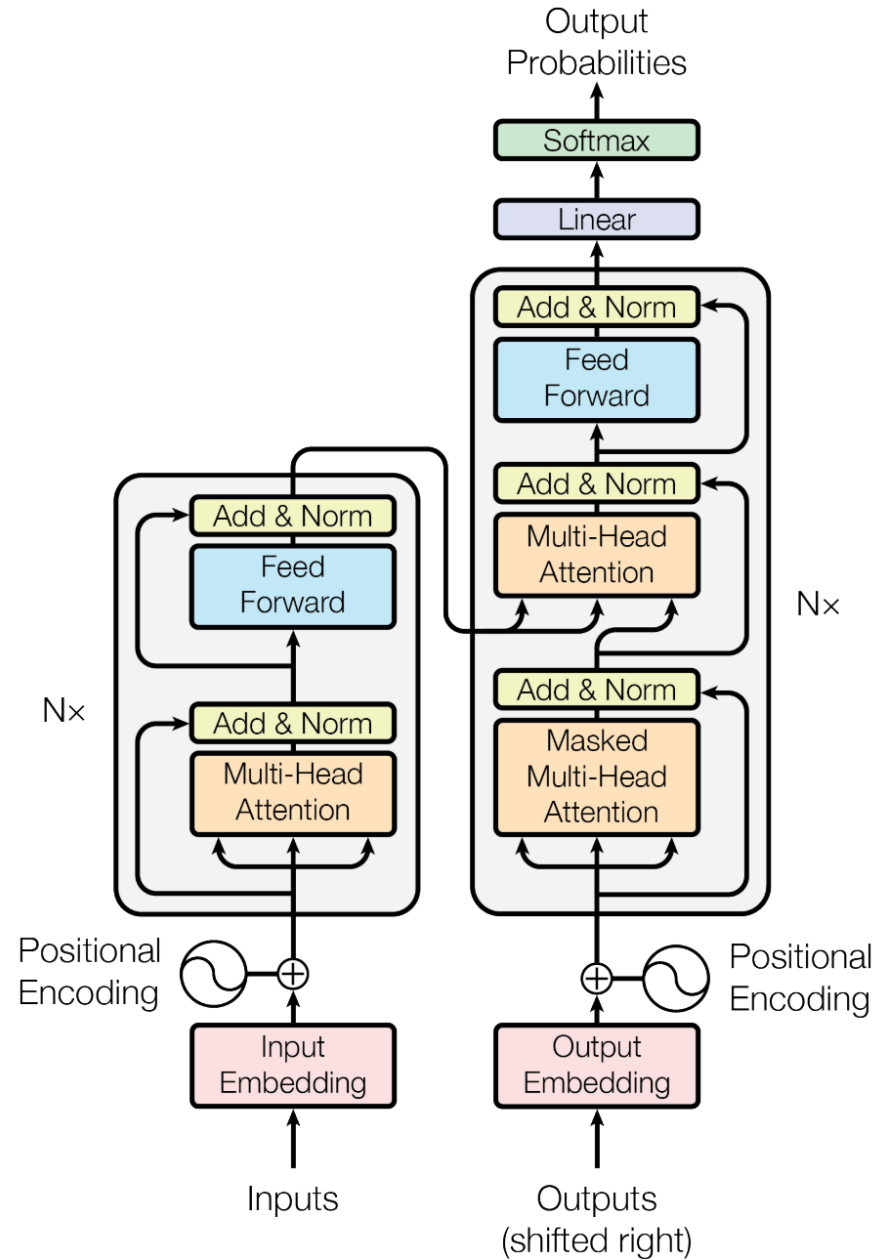
- **Non-textualized** representation
 - word-level: **word2vec**, **GloVe**, etc.
 - char-level: **fastText**, etc.
- **Contextual** representation
 - RNN: **ELMo**, etc.
 - Transformer: **GPT**, **BERT**, etc.

MTL for NLP



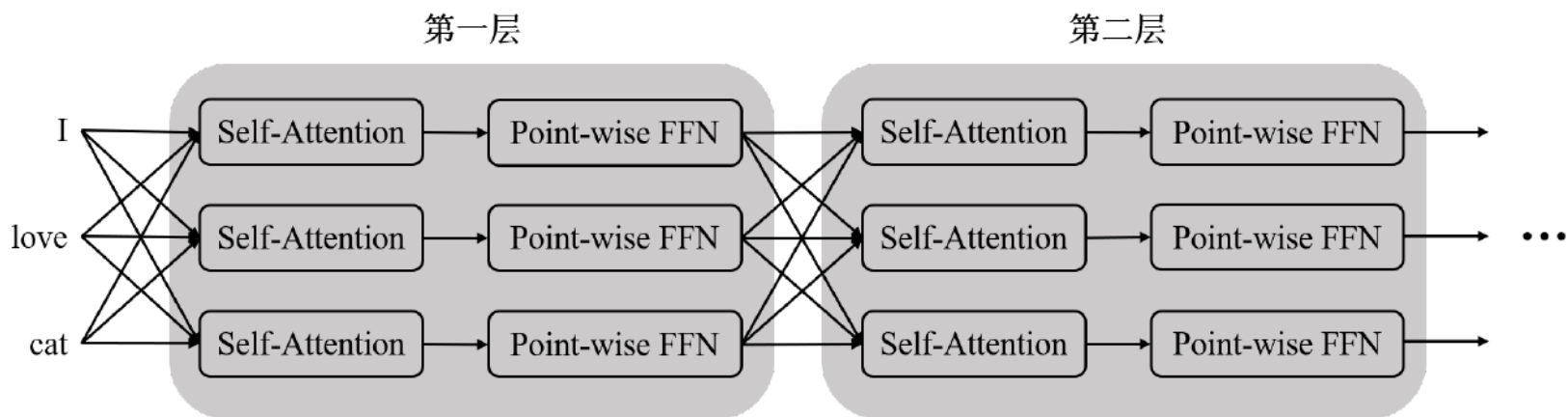
- MTL with CNNs
 - Collobert et al. A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning
 - Misra et al. Cross-stitch Networks for Multi-task Learning
- MTL with RNNs
 - Dong et al. Multi-Task Learning for Multiple Language Translation
 - Liu et al. Recurrent Neural Network for Text Classification with Multi-Task Learning
 - Hashimoto et al. A Joint Many-Task Model: Growing a Neural Network for Multiple NLP Tasks
- MTL with Transformer
 - ?

Transformer

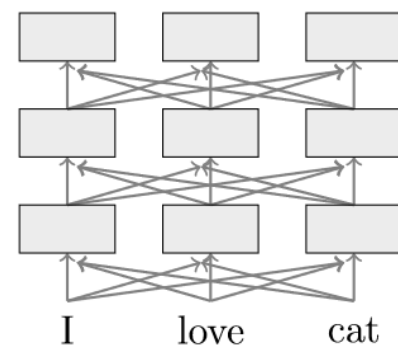


Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems*. 2017.

Transformer



Transformer编码过程



一个简化版本

MTL with Transformer

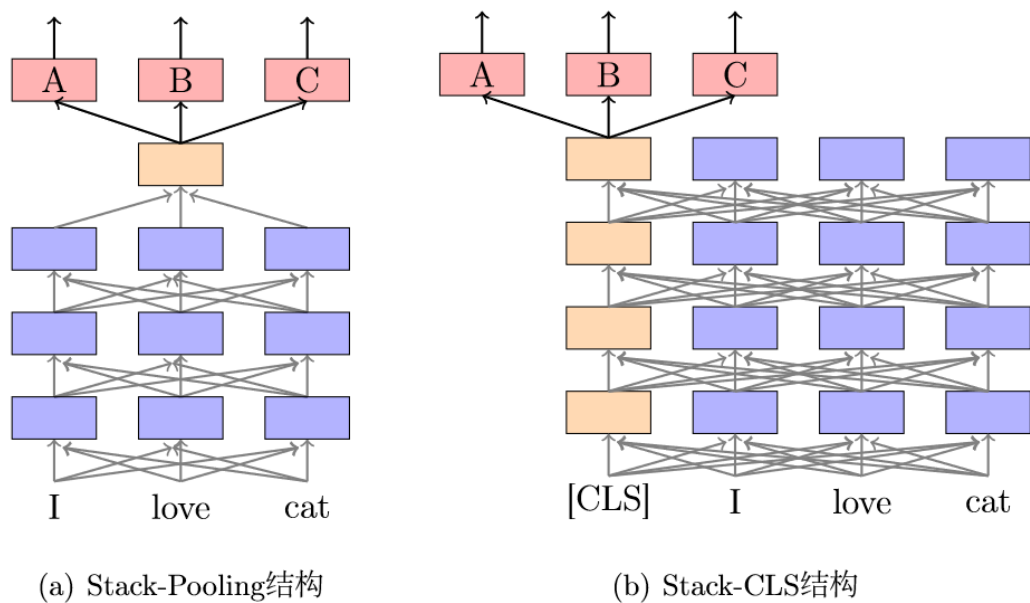


Figure 1.5: 基于Transformer的两种顶层分化共享模式

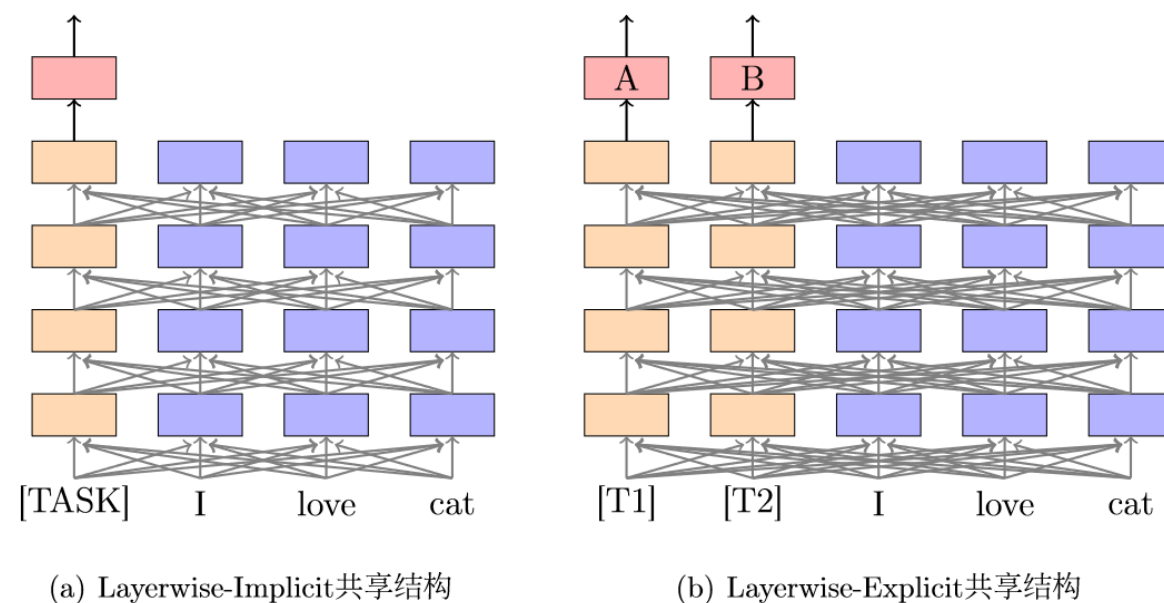


Figure 1.6: 基于Transformer的两种逐层分化共享模式

Experiments



数据集	训练集大小	验证集大小	测试集大小	类别数	平均长度	词表大小
Books	1400	200	400	2	159	19K
Elec	1398	200	400	2	101	11K
DVD	1400	200	400	2	173	20K
Kitchen	1400	200	400	2	89	9K
Apparel	1400	200	400	2	57	7K
Camera	1397	200	400	2	130	9K
Health	1400	200	400	2	81	9K
Music	1400	200	400	2	136	17K
Toys	1400	200	400	2	90	10K
Video	1400	200	400	2	156	17K
Baby	1300	200	400	2	104	8K
Mag	1370	200	400	2	117	11K
Soft	1315	200	400	2	129	11K
Sports	1400	200	400	2	94	10K
IMDB	1400	200	400	2	269	25K
MR	1400	200	400	2	21	7K

Experiments



数据集	单任务	多任务			
		S-P 结构	S-C 结构	L-I 结构	L-E 结构
Books	83.50	82.50	84.00	85.00	84.50
Elec	79.50	82.50	83.50	84.75	85.75
DVD	82.75	84.50	85.50	85.75	85.75
Kitchen	79.50	83.50	85.00	89.00	87.75
Apparel	82.75	85.50	86.75	86.00	85.75
Camera	81.75	84.25	85.00	87.00	89.00
Health	86.00	85.50	87.50	88.00	86.75
Music	76.50	83.00	83.00	82.75	81.50
Toys	80.00	84.75	86.25	88.25	86.50
Video	84.75	81.25	85.50	86.50	84.25
Baby	81.00	87.75	85.50	87.25	87.50
Mag	89.00	85.00	91.00	89.75	89.25
Soft	86.50	86.00	88.75	86.50	87.75
Sports	80.25	84.25	83.75	86.00	85.50
IMDB	80.75	84.75	85.00	84.50	84.50
MR	75.25	76.00	75.75	78.00	76.50
AVG.	81.86	83.81	85.11	85.94	85.53



三、存在的问题与解决方案

- 存在的问题
 - 任务跨度不够大
 - 数据集规模单一
 - 未在BERT上验证



四、下一步计划

➤ 实验

- BERT (a pre-trained transformer)
- NLI (MNLI, QNLI, RTE, WNLI)

➤ 论文



西安电子科技大学
XIDIAN UNIVERSITY

Q & A