

深度学习实验一四

1190200708 熊峰

2022 年 6 月 6 日

目录

1	环境配置	3
1.1	硬件配置	3
1.2	软件配置	3
2	文本分类任务	3
2.1	词嵌入	3
2.2	网络结构	3
2.2.1	RNN	4
2.2.2	GRU	4
2.2.3	LSTM	5
2.2.4	BiLSTM	5
2.3	训练过程	6
2.3.1	RNN 训练过程	6
2.3.2	GRU 训练过程	6
2.3.3	LSTM 训练过程	7
2.3.4	BiLSTM 训练过程	7
2.4	实验结果	7
2.4.1	RNN 实验结果	7
2.4.2	GRU 实验结果	8
2.4.3	LSTM 实验结果	9
2.4.4	BiLSTM 实验结果	10
2.4.5	实验结果对比	11
3	天气预测	12
3.1	数据预处理	12
3.2	网络结构	12
3.3	实验效果	12

1 环境配置

1.1 硬件配置

CPU : Intel(R) Xeon(R) Silver 4214
GPU : TITAN RTX 24G
MEM : 128G RAM

1.2 软件配置

OS : Ubuntu 20.04.1 LTS
PyTorch : Stable 1.11.0 CUDA 11.3
IDE : PyCharm 2021.3.2

2 文本分类任务

本部分主要描述模型的结构以及相关代码的编写。

2.1 词嵌入

本部分使用 FastText 训练词嵌入的模型，将词嵌入的维度设置为 300.

FastText 认为单词的词法结构会携带有关单词含义的重要信息，而传统的单词嵌入并不会考虑这些信息，传统的单词嵌入会为每个单词训练一个唯一的单词嵌入。这对于形态丰富的语言非常重要，在这种语言中，单个单词可能具有大量的形态形式，每种形态形式很少出现，因此很难训练良好的词嵌入。

FastText 尝试通过将每个单词视为其子单词的集合来解决此问题。为了简单和独立于语言，将子词视为该词的字符 n -gram (n 元)。一个单词的向量被简单地认为是其组成特征图的所有向量之和。

与原始 Word2Vec 相比，FastText 在语法任务上的表现要好得多，尤其是在训练语料库较小的情况下。

2.2 网络结构

本部分主要描述所使用的网络结构。

2.2.1 RNN

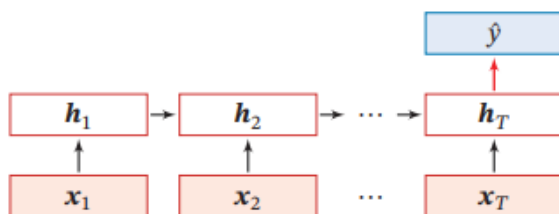


图 1: RNN Architecture

由于为文本分类任务，使用的 RNN 的结构为序列到类别的模式。

输入的样本 $x_{1:T} = (x_1, \dots, x_T)$ 为一个长度为 T 的序列，输出类别 $y \in \{0, \dots, 9\}$ ，将样本 x ，按不同时刻输入到循环神经网络中，得到不同时刻的隐藏状态 h_1, \dots, h_T ，可以将 h_T 看作整个序列的最终表示，并输入给分类器 $g(\cdot)$ 进行分类。

如上图 [1] 所示，除了使用最后时刻的状态作为整个序列的表示，也可以使用整个状态的所有状态进行平均，并使用这个平均状态作为整个序列的表示，即 $\hat{y} = g(\frac{1}{T} \sum_{t=1}^T h_t)$

2.2.2 GRU

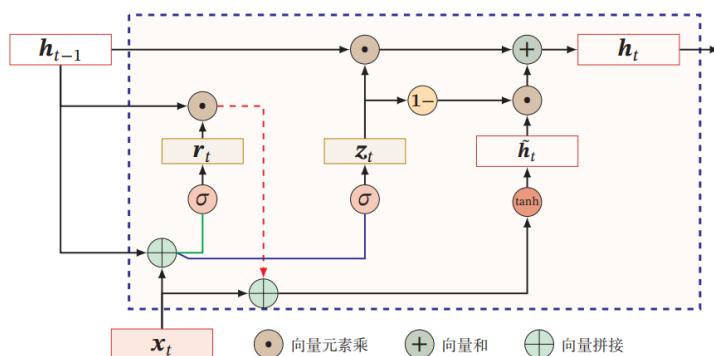


图 6.8 GRU 网络的循环单元结构

图 2: GRU Architecture

GRU 的结构如上图所示，它是一种比 LSTM 网络更加简单的循环神经网络，引入门控机制来控制信息更新的方式。

和 LSTM 不同，GRU 不引入额外的记忆单元，GRU 网络引入一个更新门来控制当前状态需要从历史状态中保留多少信息，以及需要从候选状态中接受多少新信息。

2.2.3 LSTM

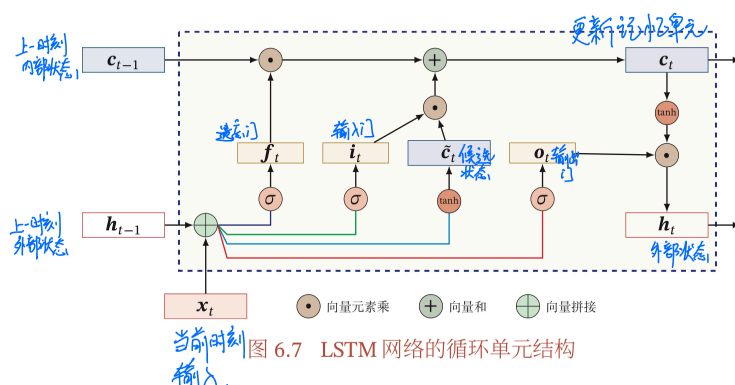


图 3: LSTM Architecture

LSTM 的结构如上图所示，其为了解决 RNN 中梯度消失的问题，引入了门控机制，来控制信息传递的路径。LSTM 中三个“门”分别为输入门 i_t 、遗忘门 f_t 和输出门 o_t 。

输入门 i_t 控制当前时刻的候选状态 \tilde{c}_t 有多少信息需要保存。

遗忘门 f_t 控制上一个时刻的内部状态 c_t 需要遗忘多少信息。

输出门 o_t 控制当前时刻的内部状态 c_t 有多少信息需要输出给外部状态。

通过 LSTM 循环单元，整个网络可以建立较长举例的时序依赖关系。

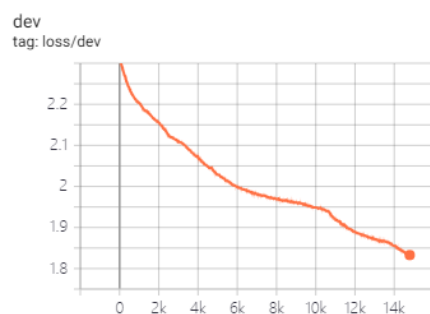
2.2.4 BiLSTM

双向 LSTM 的结构与 LSTM 的结构类似，相较于 LSTM 的主要改进是加上了反向累积的信息。

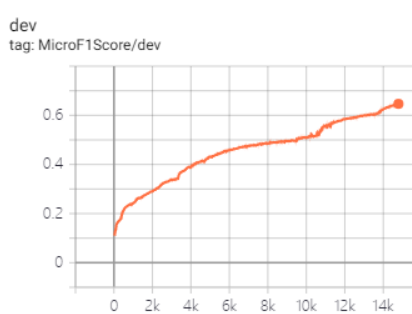
2.3 训练过程

实验中 RNN 收敛的最慢，迭代次数超过 14000 次仍未收敛，GRU、LSTM、BiLSTM 收敛速度较快。

2.3.1 RNN 训练过程



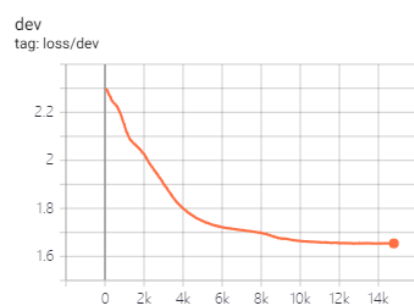
(a) RNN Train Loss



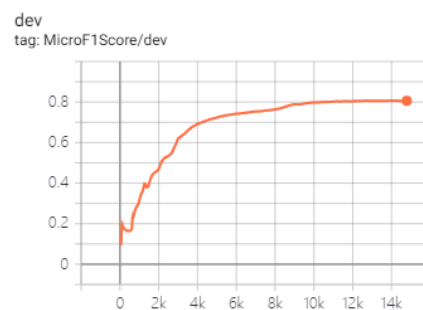
(b) RNN Train Micro F1 Score

图 4: RNN Train

2.3.2 GRU 训练过程



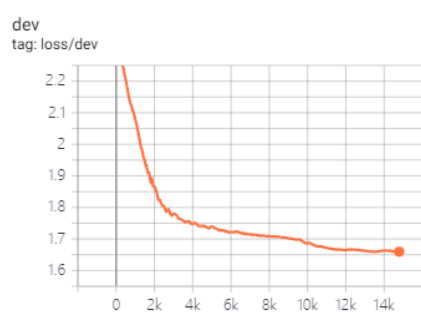
(a) GRU Train Loss



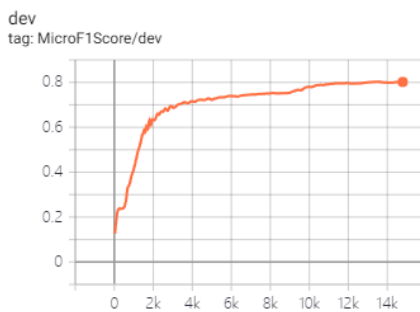
(b) GRU Train Micro F1 Score

图 5: GRU Train

2.3.3 LSTM 训练过程



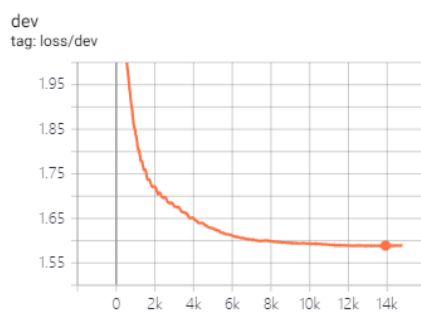
(a) LSTM Train Loss



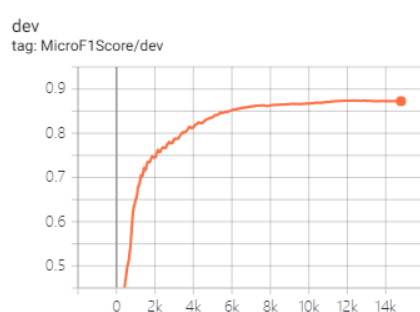
(b) LSTM Train Micro F1 Score

图 6: LSTM Train

2.3.4 BiLSTM 训练过程



(a) BiLSTM Train Loss



(b) BiLSTM Train Micro F1 Score

图 7: BiLSTM Train

2.4 实验结果

2.4.1 RNN 实验结果

RNN 的实验效果及 confusion matrix 如下所示:

class	precision	recall	f1-score	support
Book	0.69	0.76	0.72	770
Pad	0.60	0.67	0.63	2000
Phone	0.00	0.00	0.00	464
Fruit	0.55	0.61	0.58	2000
Shampoo	0.49	0.50	0.49	2000
ElectricWaterHeaters	0.00	0.00	0.00	115
Monmilk	0.52	0.60	0.56	407
Clothe	0.75	0.63	0.69	2000
Computer	0.58	0.76	0.66	798
Hotel	0.92	0.94	0.93	2000
Accuracy			0.65	12554
macro avg	0.51	0.55	0.53	12554
weighted avg	0.62	0.65	0.63	12554

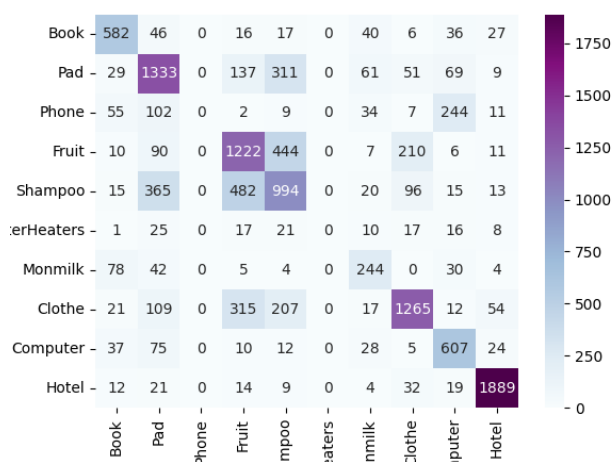


图 8: RNN confusion matrix

2.4.2 GRU 实验结果

GRU 的实验效果及 confusion matrix 如下所示:

class	precision	recall	f1-score	support
Book	0.79	0.95	0.86	770
Pad	0.53	0.82	0.64	2000
Phone	0.00	0.00	0.00	464
Fruit	0.85	0.86	0.85	2000
Shampoo	0.75	0.84	0.79	2000
ElectricWaterHeaters	0.00	0.00	0.00	115
Monmilk	0.00	0.00	0.00	407
Clothe	0.81	0.89	0.85	2000
Computer	0.00	0.00	0.00	798
Hotel	0.93	0.97	0.95	2000
Accuracy			0.76	12554
macro avg	0.47	0.53	0.49	12554
weighted avg	0.67	0.76	0.70	12554

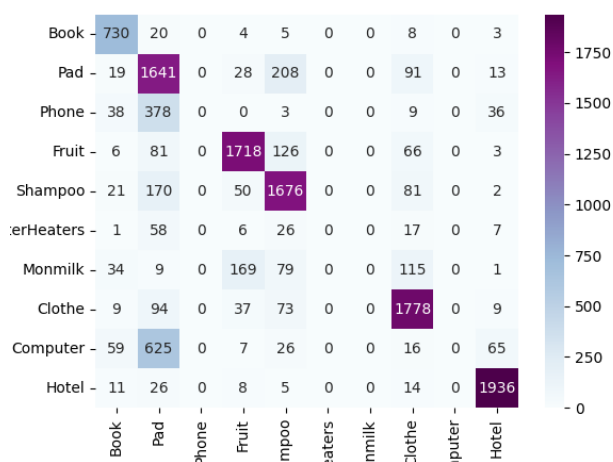


图 9: GRU confusion matrix

2.4.3 LSTM 实验结果

LSTM 的实验效果及 confusion matrix 如下所示:

class	precision	recall	f1-score	support
Book	0.59	0.93	0.72	770
Pad	0.56	0.75	0.64	2000
Phone	0.00	0.00	0.00	464
Fruit	0.91	0.84	0.88	2000
Shampoo	0.65	0.86	0.74	2000
ElectricWaterHeaters	0.00	0.00	0.00	115
Monmilk	0.00	0.00	0.00	407
Clothe	0.86	0.86	0.86	2000
Computer	0.00	0.00	0.00	798
Hotel	0.90	0.97	0.93	2000
Accuracy			0.74	12554
macro avg	0.45	0.52	0.48	12554
weighted avg	0.65	0.74	0.69	12554

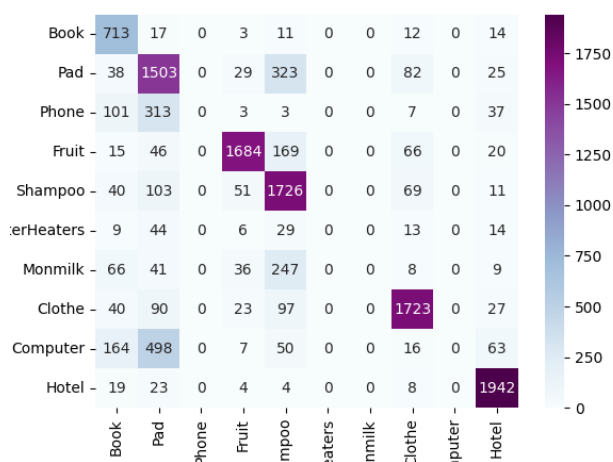


图 10: LSTM confusion matrix

2.4.4 BiLSTM 实验结果

BiLSTM 的实验效果及 confusion matrix 如下所示:

class	precision	recall	f1-score	support
Book	0.94	0.90	0.92	770
Pad	0.78	0.79	0.78	2000
Phone	0.76	0.82	0.79	464
Fruit	0.90	0.90	0.90	2000
Shampoo	0.80	0.83	0.81	2000
ElectricWaterHeaters	0.00	0.00	0.00	115
Monmilk	0.96	0.95	0.96	407
Clothe	0.87	0.88	0.88	2000
Computer	0.84	0.85	0.84	798
Hotel	0.97	0.96	0.97	2000
Accuracy			0.87	12554
macro avg	0.78	0.79	0.79	12554
weighted avg	0.86	0.87	0.86	12554

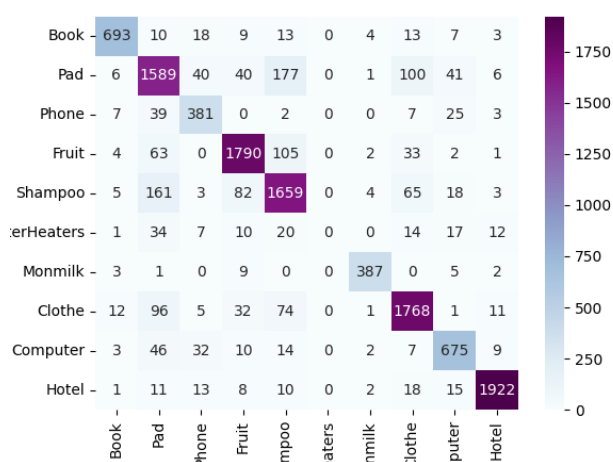


图 11: BiLSTM confusion matrix

2.4.5 实验结果对比

本次实验中 BiLSTM 的效果最好，其相较于 LSTM 累积了双向的文本的信息，相较于 RNN 引入了门控机制。LSTM 和 GRU 的效果相似。RNN

效果相较于以上模型效果较差一些，此外，RNN 收敛的速度也远远慢于上述模型。

观察实验效果表，可以看到有些类没有被分出来，一种是比较类似的情况，例如手机、电脑、平板等，可能存在评论类似等情况，一种是样本数量较少时，例如热水器、蒙牛等。

在将 GRU 和 LSTM 换成 BiLSTM 后，模型的识别能力大大增强，仅有热水器由于样本量太少，导致模型效果学习不佳。

3 天气预测

3.1 数据预处理

首先对时间的数据进行预处理，直接传入日期会令网络难以学习，进行正余弦处理后使数据之间更加贴近，例如 1 月 12 月之间的气候应该使比较接近的，使用正余弦处理可以实现这种效果。

将数据分为训练集和测试集，训练集为 2009-2014 年的数据集，测试集为 2015-2016 年的数据集，以七天为时间间隔，对数据进行分割。其中前五天的数据信息为输入的信息，后两天的温度信息作为 ground truth。

3.2 网络结构

主要使用 LSTM 网络对天气数据建模，采用异步的时序输入输出方式，采用导师训练过程，输入前五天的信息，使网络预测后两天 288 个时序点的输出。

3.3 实验效果

本次实验记录了网络预测的结果与 ground truth 的误差，并绘制了相关图片。

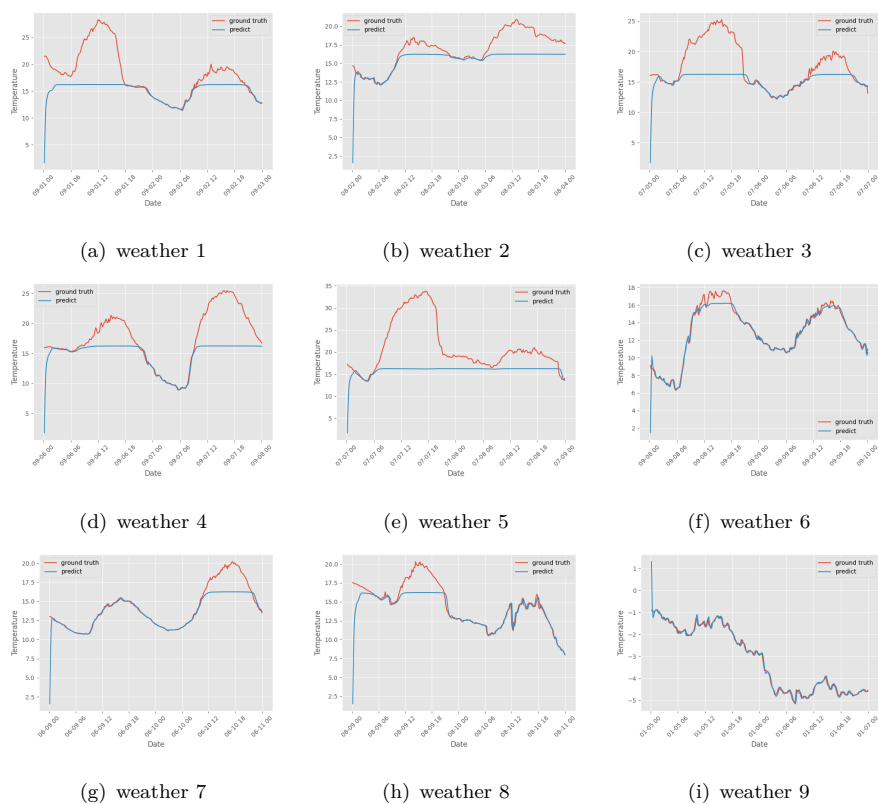


图 12: Temperature Predict

预测的平均绝对误差、绝对误差中位数、平均平方误差、平方误差中位数以及 R2Score 如下图所示，对于某些样本拟合效果较差：

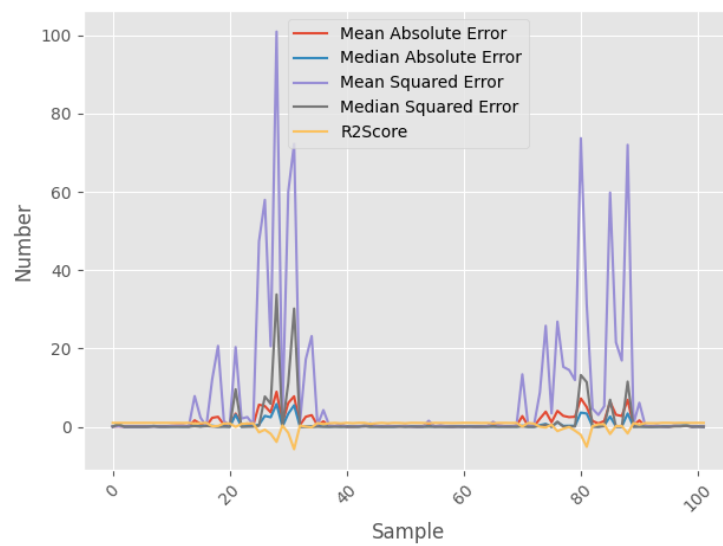


图 13: Weather Predict Error

参考文献

- [1] 邱锡鹏. 神经网络与深度学习. 机械工业出版社, 北京, 2020.