

五月份月报告

谢泽澄

这个月份主要做了以下两方面的内容，分别是期刊论文的评委回复、联机手写文本行纯 LSTM 版本的实现。

一、 期刊论文评委回复方面

- 评委质疑我们 GoogLeNet 的实验结果不够有信服力；
因此，我补充了增加 GoogLeNet 9 inception layer 的对比实验，并且在评委回复中从多个角度进行了讨论。
- 隐性语言模型中，评委并不同意我们“隐性语言模型”的假定，从 embedding 必须和字一一对应的角度否定“隐性语言模型”。
正文中我从两个方面补充了内容以反驳评委的观点：
一方面，related work 中引用在语言模型领域近期有代表性的两篇文章说明在最前沿的领域中，研究者已经开始使用 embedding 和字之间采用分布到分布的方式来完成映射。
另一方面，在第五节中，我增加了 t-SNE 图，并结合该图论述为什么我们的 embedding 和字之间使用分布到分布的映射及其意义。

二、 联机手写文本行纯 LSTM 版本的实现

目标：无约束手写实现（无约束：单字、文本行、重叠、文档、任意写）

背景：在基于 LSTM 的解决方案中，LSTM 的输入是 Δx 、 Δy 、penup\down，并没 x 和 y ，因为 x 和 y 会随着书写逐渐增大，而 Δx 和 Δy 则相对稳定，容易归一化。

任务：基于 7356 类单字合成文本，并训练纯 LSTM 版本网络，在比赛和 casia 测试集上达到满意结果。

问题：

- 归一化问题：
方案一：要求输入样本平均点间距离一致；
不合理，因为不同平台采样率不一致；
方案二：（临时方案）直接将单字放缩到 60 到 126 的范围再进行合成，这是模拟测试集的分布。
合理，但网络难以迁移到其它测试数据集。
- 字间距离问题：
各种书写方式的唯一区别：上一字结尾点和下一个字初始点之间的 Δx 和 Δy （不妨称为“字间 Δx 和 Δy ”）。
方案一：随机设定字间 Δx 和 Δy 来合成文本行；
不合理；测试结果非常低；太过相信 LSTM 的能力，不切实际。
方案二：随机使用字内的某一个 Δx 和 Δy 来合成文本行；
原理：完全模糊化切分点，让 LSTM 学习在完成字的识别后再找到切分点；
如果训练成功，则生成的 LSTM 可以适用于各种
可以接受；识别率 AR 56.68（还没有解决标点符号问题）。
方案三：在方案二训练得到的模型的基础上，加入三分之一数据字间距离随机

设定为 2~12 进行 finetune

可以接受；识别率 AR 64.07（还没有解决标点符号问题）。

● 标点符号问题：

方案一：人为放缩标点符号；

不合理；真实分布过于多样化；

方案二：加入真实数据（各占 50%）进行训练

极好；识别率 AR 91.24（还有提升空间）

识别率提升明显原因：

- 1、单字合成训练时，7356 类单字分布平均，而使用真实样本进行训练更符合文本行情况下，每个字应该出现的概率；
- 2、使用真实文本行样本训练时，标点符号的位置更加准确，真实。从而完美解决标点符号的问题。
- 3、使用真实样本训练时，字间的距离关系更加合理。

结论：

- 1、纯 LSTM 网络用于通用无约束手写识别还有诸多困难：归一化、字间距离、标点符号问题。
- 2、纯 LSTM 网络用于文本行识别任务则可以直接通过混合真实训练数据训练来解决归一化、字间距离、标点符号问题。

改进空间：（针对通用无约束手写识别）

- 1、标点符号必须设定书写规则，例如，都应该跟文本行中的标点符号的位置类似。如此可以针对性地合成样本。
- 2、字的分布问题：可以在训练后期增加常见字的出现的概率；
- 3、字间距离：需要探索