文件编号：3107-SWC2018-20180045
受控状态：■受控　□非受控
保密级别：□公司级 □部门级 ■项目级 □普通级
采纳标准：CMMI DEV V1.2

图美集

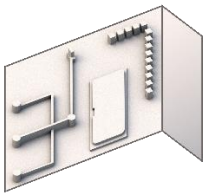
Temage

项目开发文档

Version 1.0.4

2018.11.20

Written by 3107



目录

记录更改历史

序号	更改原因	版本	作者	更改日期	备　注
1	创建	1.0.0	队员 2	2018-11-02	初步阅览文档并进行一些部分初始化
2	创建	1.0.1	队员 2	2018-11-05	撰写相关工作和技术方案
3	更新	1.0.2	队员 2	2018-11-10	撰写浏览推荐部分文档
4	更新	1.0.3	队员 2	2018-11-13	强调突出前后端工作

1 引言

1.1 编写目的

该项目技术研究报告的编写目的是为了全面深入分析和介绍本次项目的技术细节从项目的背景，到项目的整体框架设计，以及最终的实现细节，我们不断深入，层次分明的展现项目技术全貌。

该技术开发文档重点介绍了项目的技术架构和技术细节，对本项目使用的模型进行详细的阐述，对用于训练的数据集进行说明。

1.2 项目概述

我们使用 tensorflow、keras 等深度学习框架在后端进行推断和 tesorflow.js 和 keras.js 在前端进行推断，合理安排模型的分布，在本地浏览器中放置模型以承担一定量计算任务，减少服务端的运转负载及降低网络延迟，对于需要大量知识库和语料库且模型较大的功能，我们使用服务端进行推断。

1.3 项目背景

随着互联网时代的到来，互联网媒体逐渐抢占传统媒体市场，尤其是近几年的自媒体的崛起，使得传播主体多样化，平民化，普泛化。现在的网络用户只需要实名认证就可以在微博，微信公众号等自媒体平台上展现自我。因此，图文结合的使用领域越来越多，自媒体运营者为了吸引更多的用户，在文章内容和形式上绞尽脑汁。本项目希望设计一款使用深度学习技术的 web 应用，为用户提供个性化的图文结合和文本编辑服务，并以长图或其他格式发布到各大平台。

图片识别在近些年有巨大的发展，在 ILSVRC 2012 中，Alex Krizhevsky 基于 GPU 实现了有 60million 参数的模型——AlexNet，赢得了比赛的第一名。这个工作是开创性的，它引领了接下来 ILSVRC 的风潮。随后几年中，Google，Baidu 等大公司也加入到其中，得到了错误率更低的模型。同时，深度学习在自然语言处理中也大展身手。2013 年，Google 开源了一款用于词向量计算的工具——word2vec，引起了工业界和学术界的关注。随后提出的 RNN，LSTM 更是大展身手，TextCNN 在情感分析等方面更是有着令人惊叹的效果。

在本项目中，我们将结合深度学习中图片识别和自然语言处理这两个部分，为用户提供具有优良性能的图文结合、智能排版编辑功能。

1.4 术语和缩略语

[1] Tensorflow： Tensorflow 是一个采用数据流图(data flow graphs),用于数值计算的开源软件库。

[2] Keras：Keras 是一个高层神经网络 API，Keras 由纯 Python 编写而成并基 Tensorflow、

Theano 以及 CNTK 后端。

[3] Tornado: Tornado 是一种 Web 服务器软件的开源版本。Tornado 和现在的主流 Web 服务 器框架（包括大多数 Python 的框架）有着明显的区别：它是非阻塞式服务器，而且速度相 当快。

[3] Django: Django 是 Python 编程语言驱动的一个开源模型-视图-控制器（MVC）风格的 Web 应用程序框架。

1.5 参考资料

[1] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[2] Peters M E, Neumann M, Iyyer M, et al. Deep contextualized word representations[J]. arXiv preprint arXiv:1802.05365, 2018.

[3] Lai S, Xu L, Liu K, et al. Recurrent Convolutional Neural Networks for Text Classification[C]//AAAI. 2015, 333: 2267-2273.

[4] Zeman D, Hajič J, Popel M, et al. CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies[J]. Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, 2018: 1-21.

[5] Fares M, Kutuzov A, Oepen S, et al. Word vectors, reuse, and replicability: Towards a community repository of large-text resources[C]//Proceedings of the 21st Nordic Conference on Computational Linguistics, NoDaLiDa, 22-24 May 2017, Gothenburg, Sweden. Linköping University Electronic Press, 2017 (131): 271-276.

2 问题聚焦

2.1 问题描述

基于设想，我们有三个主要问题需要解决：

1. 如何根据用户输入给予用户风格推荐
2. 如何将用户的图片和文字进行匹配并合理排版
3. 如何推荐相关内容使得用户可以从中得到借鉴

2.2 问题抽象

根据问题描述提出的问题，我们进一步抽象问题

针对问题一，我们既需要从用户的文本中提取特征，又需要根据用户使用历史，向用户

提供风格进行选择。

针对问题二，我们需要从图像和文本中分别提取特征，并将两者进行比较，从而进行匹配，通过匹配进行文本和图片的融合。可以抽象为三个部分，第一部分是将图片映射到向量空间，第二部分是将文字映射到向量空间，第三部分是在这个向量空间中分析文本向量和图片向量的相似度，以该相似度进行匹配。

针对问题三，我们需要对用户作品进行分类，在用户进行搜索时给用户提供相关作品的展示。

2.3 问题定位

本项目中的技术问题主要为自然语言处理和图像识别方面的问题。

2.4 问题评估

对于问题一，文本的特征提取在自然语言处理领域中，是较为成熟的一个部分，可供选择的模型较多，普适性高；另一方面，基于时序性操作的推断，可以使用 LSTM 神经网络模型。

对于问题二，图像识别在计算机视觉领域也是较为成熟的一个部分，众多的团队提供了非常多深度学习的模型可供挑选和 fine-tuning，文本嵌入也是近几年提出的风靡全球的模型，从 word2vec 到 doc2vec 以及最新的 ELMo，结合 RNN、LSTM，我们有成熟的方案能够解决这个问题。

对于问题三，我们对比相似性的依据应当是文本中的关键词和问题二中推断出的风格主题。对于关键词的提取，一些基于统计的方法（TF-IDF）可以非常好地达到效果，我们使用关键词匹配来进行推荐。

2.5 问题分解

问题一可分解为文本分类和用户习惯追踪两个子问题。文本分类问题，可以使用较为成熟的模型进行训练，推断。用户习惯追踪我们可以使用 LSTM 神经网络模型。结合两个部分，给予用户最终的风格推荐。

问题二较为复杂，可分解为图像识别，文本嵌入和图文匹配三个问题。图像识别问题可以使用是使深度学习在众多机器学习算法中脱颖而出的 CNN 模型，基于 CNN 开发的模型种类繁多，可供本项目进行挑选和 fine-tuning。文本嵌入问题可以使用 RNN-LSTM 对文中单词或句子进行 encode，得到表示单词或句子的向量。对于图文匹配问题，我们可以基于余弦计算等方法找到找到最为匹配的图片与文字，再使用基于统计的方法，对文章进行排版。

问题三有成熟的解决方案，我们对作品文本提取关键词，进行比对，为提高搜索效率，我们可以使用 sphinx 开源搜索引擎框架对数据库建立索引。

3 相关工作

1. VGG 卷积神经网络是牛津大学在 2014 年提出来的模型。当这个模型被提出时，由于它的简洁性和实用性，马上成为了当时最流行的卷积神经网络模型。它在图像分类和目标检测任务中都表现出非常好的结果。在 2014 年的 ILSVRC 比赛中，VGG 在 Top-5 中取得了 92.3%的正确率。
2. ELMo 于 2018 年 2 月由 AllenNLP 提出，与 word2vec 或 GloVe 不同的是其动态词向量的思想，其本质即通过训练 language model，对于一句话进入到 language model 获得不同的词向量。根据实验可得，使用了 ELMo 词向量之后，许多 NLP 任务都有了大幅的提高。
3. 2015 年，中科院在 AAAI 上发表了一篇名为《Recurrent Convolutional Neural Networks for Text Classification》的论文，并提出了 Recurrent Convolutional Neural Network(RCNN)的模型用于在O(n)时间内进行文本分类。在论文提及的数据集中效果均优于原有的分类模型。

4 技术方案

4.1 技术方向

本项目中将对文本进行处理，所使用的模型中将会用到 RNN(LSTM)、CNN。

4.2 模型选择

4.2.1 模型设计

对于文本分类问题，我们决定使用 ELMo 对单词进行 embedding，之后使用 RNN(LSTM)对文章进行 embedding，之后使用 TextCNN 进行文本分类。

对于用户习惯追踪，我们使用成熟的 LSTM 神经网络建立模型进行训练。

对于图像识别问题，我们使用 CNN 对图像进行卷积，输出向量，将向量和 ELMo 模型产出的单词 embedding 进行比对，得到最佳匹配。

为分担服务器工作，提高用户体验感，我们将文本分类所使用的 TextCNN 模型放置在用户前端，并使用 tensorflow.js 等前端深度学习框架让模型在前端进行推断。

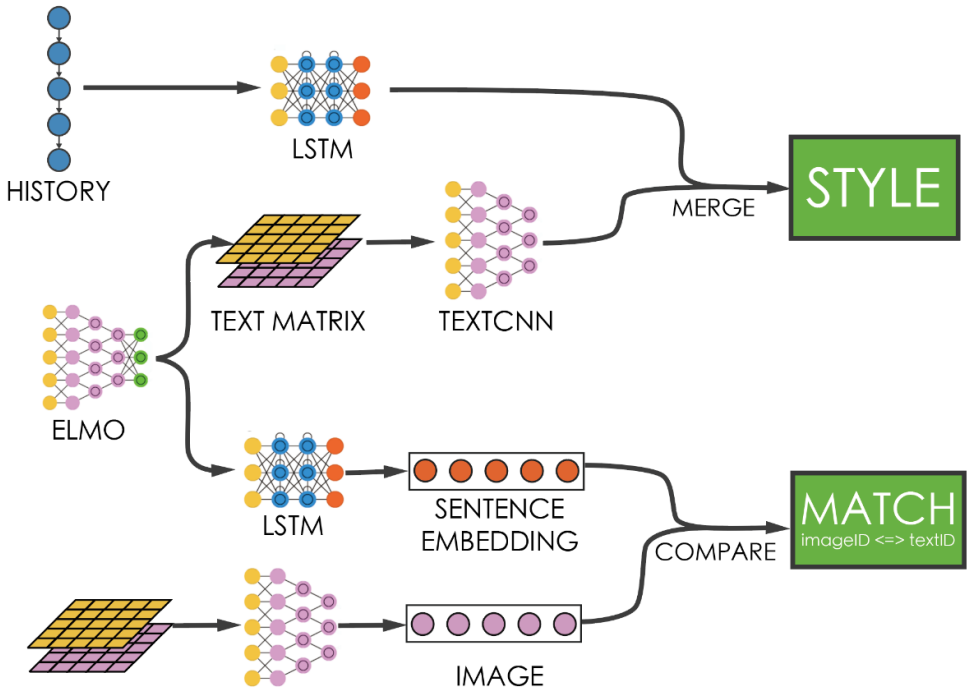


图 4.1 模型设计图

4.2.2 模型结构

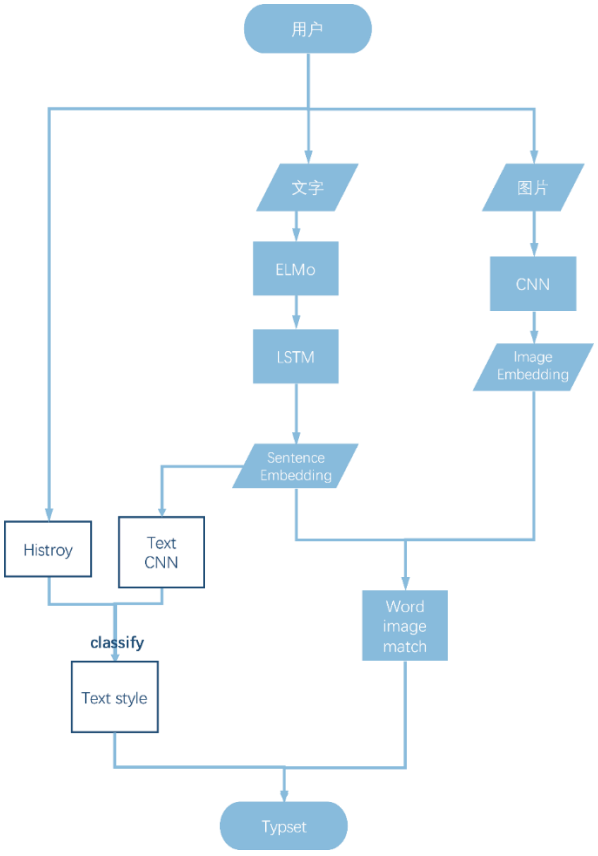


图 1.1 模型流程图，其中 x 色代表用户浏览器中执行的流程，x 色代表服务端执行的流程

4.2.3 数据集

1. 文本分类模型： 今日头条数据集，共 382688 条，分布于 15 个分类中。用于文本分类进行风格推荐的模型训练，采集时间为 2018 年 5 月，凭借此我们将得到新闻中的图片和图片上下文，上述数据可用于文本分类的训练和图文匹配的训练。
2. 用户习惯追踪模型：使用爬虫在今日头条上爬取用户的文章发布历史，使用训练好的文本分类模型对文章进行文本分类，将结果用于用户习惯追踪模型的训练。
3. 图像识别模型：今日头条数据集，共 382688 条，凭借此我们将得到新闻中的图片和图片上下文，从图片上下文结合 ELMo 模型，得到与图片对应的向量嵌入。
4. ELMo 模型我们准备 fine-tuning HIT-SCIR/ELMoForManyLangs

上述模型在训练过程中训练集(train)、验证集(develop)，测试集(test)按照 7:2:1 进行随机划分。

4.3 结果期望

模型在推断后，图片与文本结合合理，页面美化与文本风格统一，用户仅需要简单的修改就可生成最终作品用于各大平台发布。我们希望项目上线初期用户的评分反馈能在 4 星及以上。