

Abstract

职位收入对于求职者和人力资源平台都是极为重要的信息。通过职位描述实现对职位收入的预测，将有助于提高求职者寻找到合适工作的效率。

在本次工作中，我们使用深度学习模型对职位描述信息进行分析，对职位的收入进行预测。

此次使用的双向门控循环单元卷积神经网络模型较比之前使用的文本卷积模型，在精度上实现了16.5%的提升。此外，我们还对其他模型进行了实验对照。

关键词：收入预测，深度学习，GRU，CNN

Introduction

求职信息的不对等一直使求职者处于相对不利的状态，特别是职位收入信息，一般情况很难获取。通过深度学习技术对上百万的求职信息进行分析，可以实现对职位收入的有效预测。

职位信息的数据来源于STANBY职位搜索引擎，实验中使用的训练数据包含超过八百万条求职信息。双向门控循环单元卷积神经网络模型在实验中的表现也超过了一般TextCNN, RCNN, Bid-LSTM, ResNet 等模型。

Model

TextCNN

由Kim提出的多通道TextCNN模型在精度和训练速度上都有很好的表现。(Kim, 2014)

但是CNN无法建立长距离依赖关系的特点，影响模型结合文本上下文内容进行语义的表征学习。

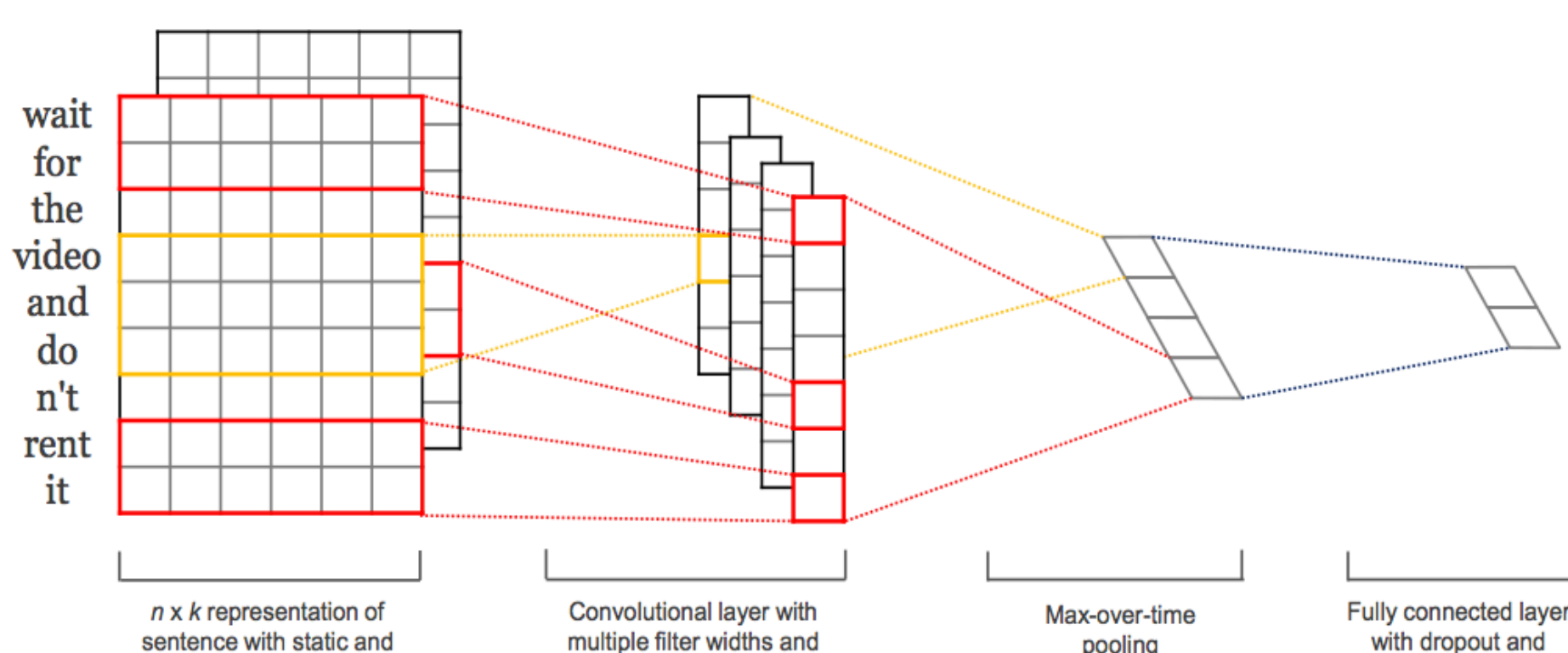


Figure 1. TextCNN (Kim, 2014)

Model

Bidirectional-GRU-CNN

在CNN进行表征学习之前，通过双向门控循环单元使每个单词可以结合上下文内容进行语义学习，弥补单纯使用CNN的不足。使用双向门控循环单元结构，可以有效缓解对较长句子学习时，产生的梯度下降问题。

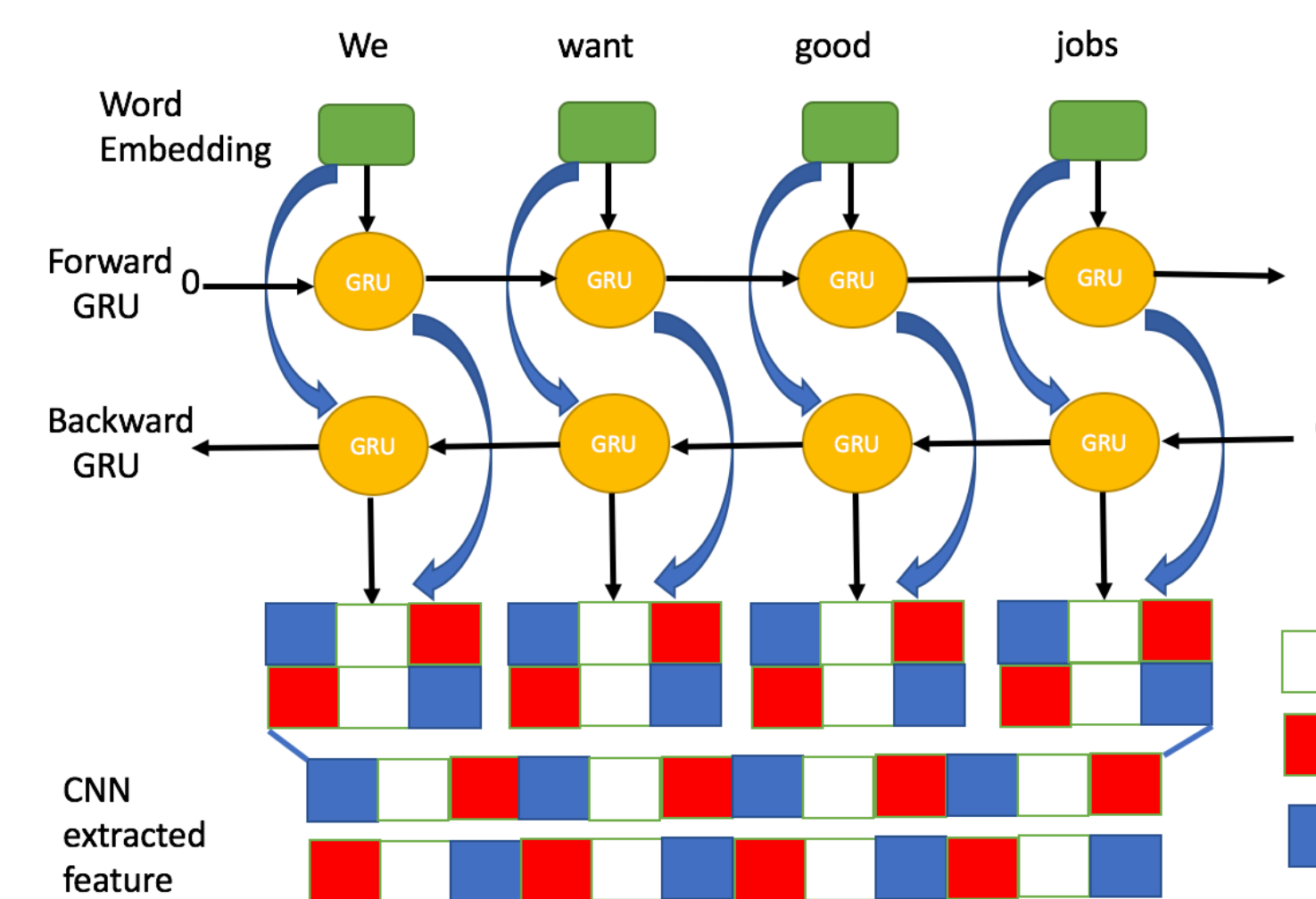


Figure 2. Bid-GRU-CNN

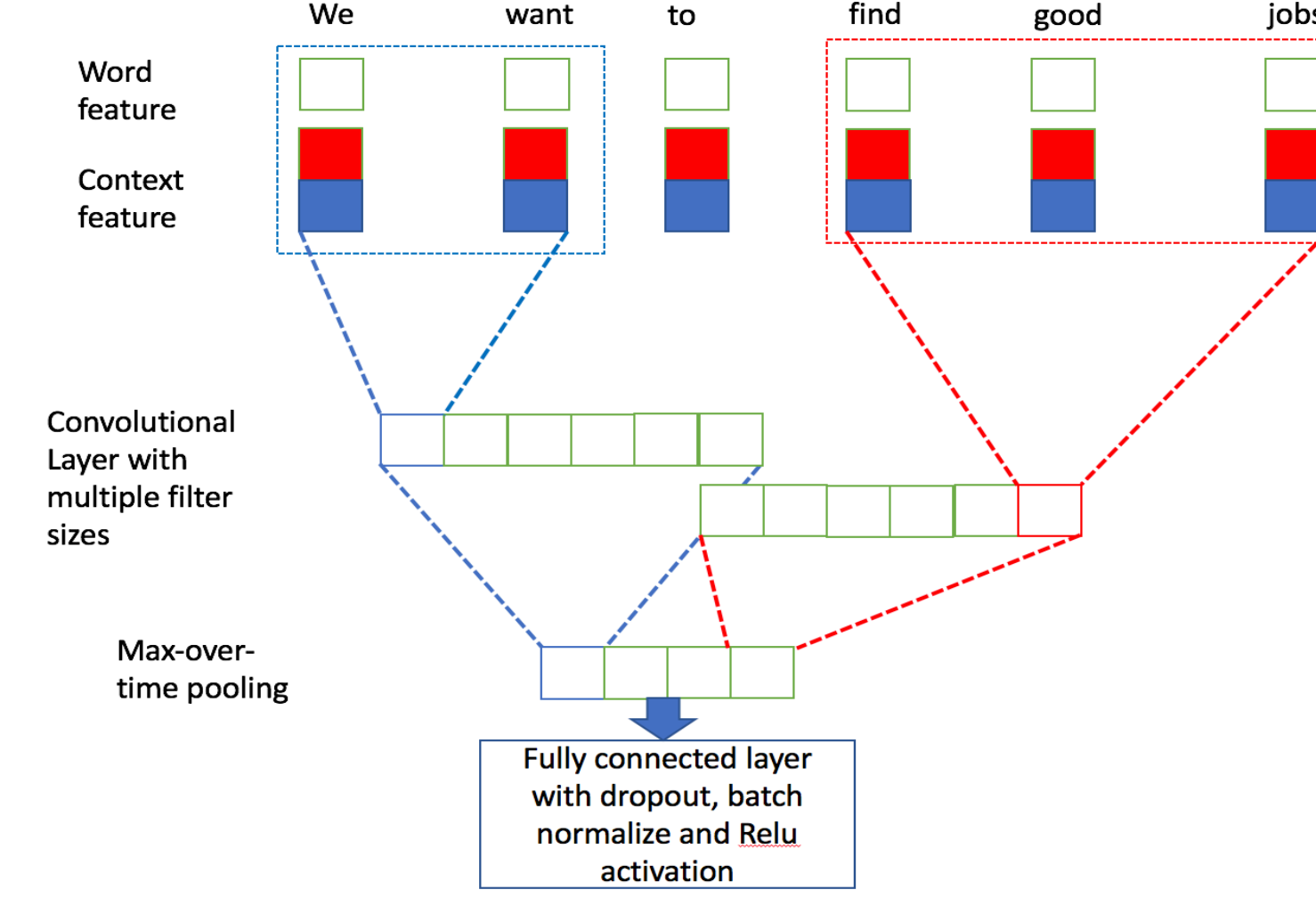


Figure 3. Multi-channel CNN.

Experiments & Results

Experiments

STANBY数据：求职信息包括 职位标题，职位描述，职位要求，工作时间，地点，工作类型，以及收入信息。

英伟达GPU：训练数据包含超过八百万条的职位信息，模型结构中同时包含循环和卷积神经网络，英伟达GTX1080 GPU帮助我们在有限的时间内，实现对模型的训练以及改进。

Pre-trained word embedding：使用FastText模型和STANBY进行预训练(200-dim, 34million tokens)

Results

我们将职位信息分成四个等级进行进行结果评估：

1, 年收少于36万 2, 年收入少于48万 3, 年收入少于60万 4, 全范围
使用平均绝对差对实验结果进行评估。



Figure 4. 模型参数

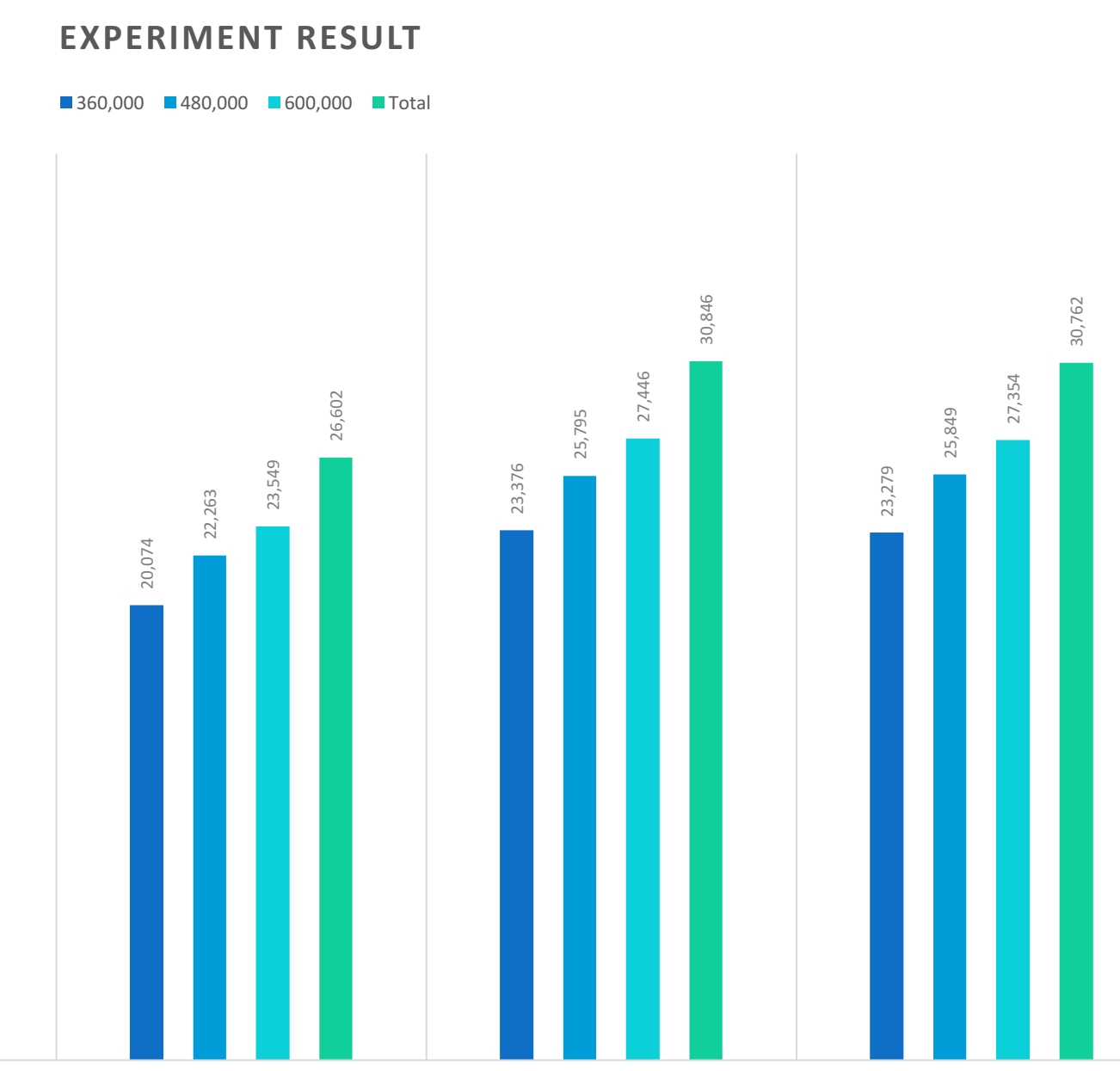
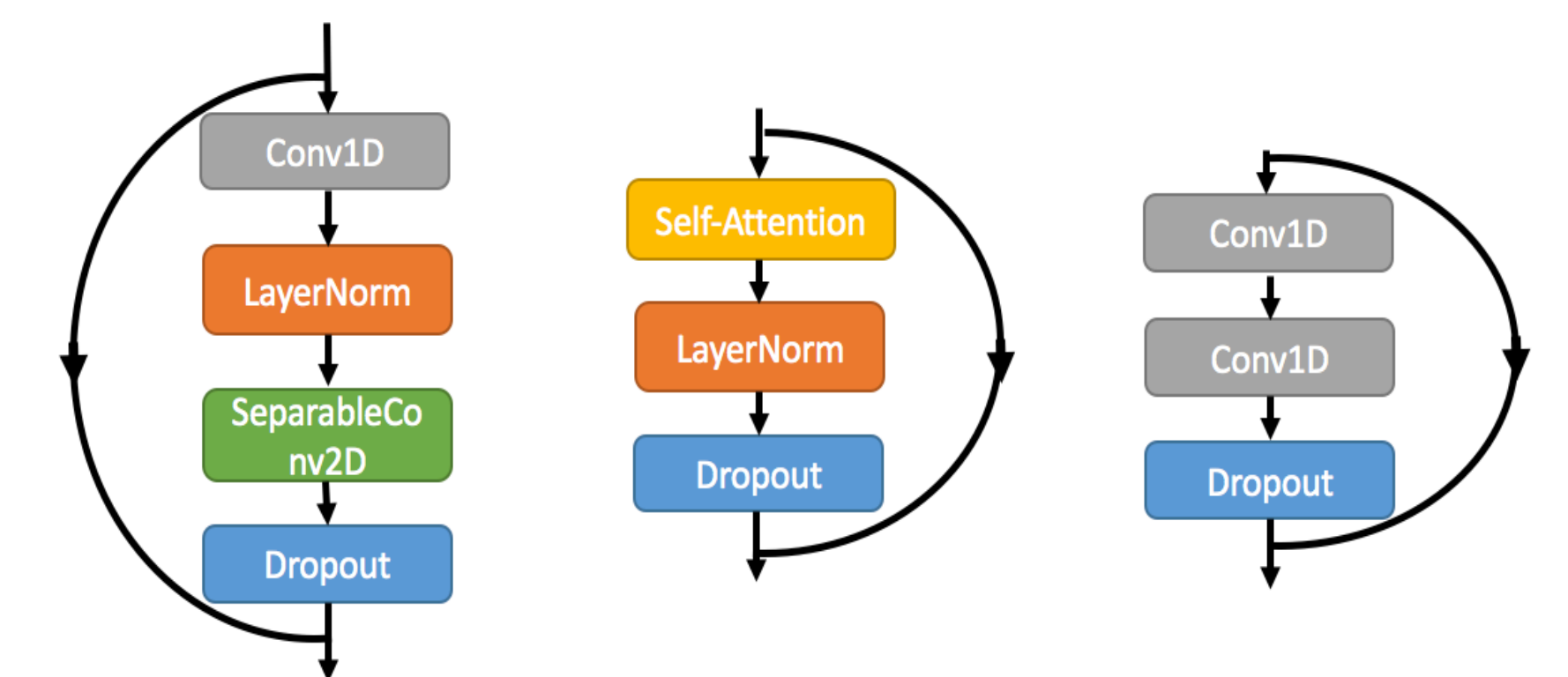


Figure 5. 预测结果

Discussion

1. ResNet并没能提升精度，可能原因采用在输入层设置ResNet。
2. GRU在精度和训练速度上都强于LSTM，可能原因在于在输出结构中，减少一个Gate。
3. 使用Raw数据，噪音产生一定影响。
4. 对其他结构进行尝试，过于复杂模型难于应用在实际业务中。



Conclusions

此次训练的双向门控循环卷积神经网络在实际业务中取得了良好的效果。

英伟达的GPU保证了项目时间得以控制在有效范围内，保证了项目模型的持续运作。

很多论文中有效高精度的方法，在实际业务中存在种种限制，简单高效的模型是将深度学习技术落地的关键。

在不同模型结构的尝试过程，也需要大量的时间进行验证。少量数据的验证往往不能足以证明模型的性能和鲁棒性。（很多论文里的数据量过小）

高效的GPU也有效减少了对多模型的验证时间。

Future Directions

今后的工作重点还是对简单高效的模型进行探索，同时，在不使用循环神经的情况下，探索GPU对大量数据进行并行处理的方法改进。

帮助更多的求职者高效的寻找合适的工作。

Contact Information

Email: zhongsheng.wang@bizreach.co.jp
Website: <https://www.bizreach.co.jp/>
<https://jp.stanby.com/>

References

1. Kim, Y. 2014. Convolutional Neural Networks for Sentence Classification. In *proceedings of EMNLP 2014*.
2. Lai, D., Xu, L., Liu, K., and Zhao, J. 2015. Recurrent convolutional neural networks for text classification. Page 2267-2273. In *proceedings of AAAI 2015*.
3. Srivastava, Rupesh Kumar; Greff, Klaus; Schmidhuber, Jurgen. May 2015. Highway Networks. arXiv:1505.00387
4. He, K., Zhang, X., Ren, S., and Sun, J. . 2016. Deep residual learning for image recognition. In *proceedings of CVPR 2016*.

Acknowledgements

感谢STANBY提供的职位信息数据。
感谢英伟达公司组织这次会议。