卷积神经网络在音频分类中的优化

1. 音频分类
   1. 什么是音频分类[1]

1.2． 音频分类的用途

情景感知计算[2]、智能可穿戴借口[3]、移动机器人导航[4]

1. 深度学习

深度学习是允许由多个处理层和多个抽样层学习数据表示的计算机模型。**这些方法可以显著提高了最先进的语音识别,视觉物体识别、目标检测和许多其他领域如药物发现和基因组学等领域. 深度学习使用反向传播（BP）算法发现了在大型数据集中的复杂结构来说明如何机器应该如何改变其内部参数用于计算每一层对于上一层的表示。**深度卷积网络在图像处理、视频、语音和音频领域带来了突破，而递归网络对于连续的数据如文本和语音有很好的性能。

**深度学习在阻碍人工智能发展的问题解决方面做出了重大进步。它被证明是非常善于发现高维数据的复杂结构，因此适用于很多科学领域、商业和政府。它打破了图像识别[5-8]、语音识别[9-11]和许多机器学习领域的记录，如预测潜在药物分子的活性[12]，分析颗粒加速数据[13,14]，重建大脑回路[15]，预测非编码基因序列突变对基因表达和疾病的影响[16,17]。或许更令人惊讶的是，深度学习在自然语言理解方面的各种课题也产生了很好的效果[18]，特别是主题分类，情感分析，自动问答[19]和语言翻译[20,21]。**

**我们认为深度学习将有更大的成功在不远的而将来，因为他需要很少的人工干预，可以很容易的利用计算量和数据的增长。目前正在为深层神经网络而开发的新的学习算法和结构只会加快这一进程。**

1. 卷积神经网络

3.1. 什么是卷积神经网络

卷积神经网络**被设计用来处理多重矩阵的数据，例如，一个彩色图像由3个2D矩阵组成。许多数据形态是多重矩阵的形式：1D信号和序列包括语音，2D图像或声谱图，3D视频或立体图像。卷积网络利用4个关键点来利用自然信号的属性：局部连接，权值共享，池化和多层结构。如图2，一个卷积网络的内部**

**卷积层的功能是检测前一层的特征的局部连接，池化层的作用是是合并特征相似的地方。这是因为形成一个主题的特征的相对位置会有所不同。一个典型的池化单元计算一个局部块的最大值，邻近的池化单元按照一行或一列或者更多的方式移动的从局部块中获取数据，从而减少了维数，增加了移动或扭曲的不变性。2到3个卷积层，非线性函数和池化层堆叠，再连接上全连接层，通过bp算法训练所有过滤器的权重，就得到了**卷积神经网络**。**

**深度神经网络开发自然信号层级组成的特性，其中，高水平的特征由低水平的组合获得。在图像中，边缘的局部组合形成图案，图案构成部分，部分组成目标。相似的结构存在于语音和文本中，如电话里的声音，音位，音节，文本里的单词和句子。当前一层的元素的位置或外貌变化时，池化能保证表示几乎不变。**

**卷积网络中卷积层和池化层是由视觉神经科学中的简单细胞和复杂细胞的经典观念启发得到的，这种细胞是以LGN–V1–V2–V4–IT这种形式形成视觉回路的。当被给予相同图片时，卷积网络可以解释猴子的下颞叶皮质的随机160个神经元的变化。卷积网络有神经认知的根源，有相似的结构，但神经认知中没有类似bp算法之类的端到端的监督学习算法。一个原始的1D卷积网络称作时延神经网络被用于音素和简单单词的识别。**

**追溯到1990s早期，卷积网络已经有了大量的应用，开始是时延神经网络用于语音识别和文本阅读。文本阅读系统使用一个训练好的卷积网络和一个而受到语言约束的概率模型的二联合。到1990s晚期，这个系统阅读美国所有支票的10%。大量基于卷积网络的视觉特征识别和手写体识别被Microsoft开发。在1990s早期，卷积网络也被实验在人脸，手写图像的目标检测和人脸识别。**

3.2. 卷积神经网络的应用

1. 为什么要用深度学习解决音频分类
2. 研究目标与内容

参考文献

1. Gerasch M C. Acoustic Scene Classification[J]. 2015.
2. B. Schilit, N. Adams, and R. Want, “Context-aware computing applications,” in Mobile Computing Systems and Applications, 1994. WMCSA 1994. First Workshop on. IEEE, 1994, pp. 85–90.
3. Y. Xu, W. J. Li, and K. K. Lee, Intelligent wearable interfaces. John Wiley & Sons, 2008.
4. LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436.
5. Krizhevsky, A., Sutskever, I. & Hinton, G. ImageNet classification with deep convolutional neural networks. In Proc. Advances in Neural Information Processing Systems 25 1090–1098 (2012).
6. Farabet, C., Couprie, C., Najman, L. & LeCun, Y. Learning hierarchical features for scene labeling. IEEE Trans. Pattern Anal. Mach. Intell. 35, 1915–1929 (2013).
7. Tompson, J., Jain, A., LeCun, Y. & Bregler, C. Joint training of a convolutional network and a graphical model for human pose estimation. In Proc. Advances in Neural Information Processing Systems 27 1799–1807 (2014).
8. Szegedy, C. et al. Going deeper with convolutions. Preprint at http://arxiv.org/ abs/1409.4842 (2014).
9. Mikolov, T., Deoras, A., Povey, D., Burget, L. & Cernocky, J. Strategies for training large scale neural network language models. In Proc. Automatic Speech Recognition and Understanding 196–201 (2011).
10. Hinton, G. et al. Deep neural networks for acoustic modeling in speech recognition. IEEE Signal Processing Magazine 29, 82–97 (2012).
11. Sainath, T., Mohamed, A.-R., Kingsbury, B. & Ramabhadran, B. Deep convolutional neural networks for LVCSR. In Proc. Acoustics, Speech and Signal Processing 8614–8618 (2013).
12. Ma, J., Sheridan, R. P., Liaw, A., Dahl, G. E. & Svetnik, V. Deep neural nets as a method for quantitative structure-activity relationships. J. Chem. Inf. Model. 55, 263–274 (2015).
13. Ciodaro, T., Deva, D., de Seixas, J. & Damazio, D. Online particle detection with neural networks based on topological calorimetry information. J. Phys. Conf. Series 368, 012030 (2012).
14. Kaggle. Higgs boson machine learning challenge. Kaggle https://www.kaggle. com/c/higgs-boson (2014).
15. Helmstaedter, M. et al. Connectomic reconstruction of the inner plexiform layer in the mouse retina. Nature 500, 168–174 (2013).
16. Leung, M. K., Xiong, H. Y., Lee, L. J. & Frey, B. J. Deep learning of the tissue-regulated splicing code. Bioinformatics 30, i121–i129 (2014).
17. Xiong, H. Y. et al. The human splicing code reveals new insights into the genetic determinants of disease. Science 347, 6218 (2015).
18. Collobert, R., et al. Natural language processing (almost) from scratch. J. Mach. Learn. Res. 12, 2493–2537 (2011).
19. Bordes, A., Chopra, S. & Weston, J. Question answering with subgraph embeddings. In Proc. Empirical Methods in Natural Language Processing http:// arxiv.org/abs/1406.3676v3 (2014).
20. Jean, S., Cho, K., Memisevic, R. & Bengio, Y. On using very large target vocabulary for neural machine translation. In Proc. ACL-IJCNLP http://arxiv.org/ abs/1412.2007 (2015).
21. Sutskever, I. Vinyals, O. & Le. Q. V. Sequence to sequence learning with neural networks. In Proc. Advances in Neural Information Processing Systems 27 3104–3112 (2014).