

# 语音识别中的稀疏深度学习方法

刘超, CSLT  
指导老师: 王东  
May.19, 2014

# 提纲

- 选题背景
- 相关工作
- 研究方案
- 工作计划
- 参考文献

# 选题背景

- 深度神经网络已在语音识别领域得到广泛应用，取得了令人瞩目的性能提升效果
- 仍存在问题
  - 噪声 / 信道鲁棒性差
  - 数据少易过拟合，陷入局部最优
  - 参数多，计算效率低
- 全连接结构，弱先验假设

# 选题背景

- 稀疏化编码
  - 人脑激发少数神经元完成特定信息表达
  - 相当于结构化先验知识
  - 提高抗噪性能，更合理的网络结构，减少运算时间

# 相关工作

- 特征编码
  - W.J. Smit, 2008
- 节点输出稀疏化
  - G.S.V.S. Sivaram, 2010( 目标函数正则化 )
  - Nitish Srivastava, 2013(Dropout)
- 权重稀疏化
  - Dong Yu, 2012( 训练后剪裁 )

# 相关工作

- 现有方法挑战
  - 更系统，深入的结合
  - 更适应语音编码特性
  - 层次化结构研究

# 研究目标

- 稀疏编码应用于深层网络结构方法
- 稀疏网络剪裁与优化策略
- 层次化稀疏网络学习

# 研究方案

- 实现相关工作中所涉及到的稀疏学习方法
- 对比不同方法的识别率与时间效率
- 尝试提出 / 改进得到更合适的稀疏学习方法
- 探索层次化稀疏网络结构

# 实验方法

- 多种应用场景，对比不同方法效果
  - 纯净语音数据库中加入噪声
  - 移动设备端
  - 小语种小数据集
- 使用 Kaldi 搭建基础实验平台，利用 GPU 集群加快速度

# 前期工作

- 基于对角化 Hessian 矩阵的网络权重裁剪
  - 相对权重幅度裁剪，更精准地裁剪不重要边权

Pruning	WER%	
	Magnitude	OBD
0%	24.04	
50%	24.01	24.14
75%	24.36	24.32
88%	25.41	25.09
94%	27.64	26.49

# 研究计划

- 2014.05 - 2014.06
  - 调研各种稀疏学习方法
- 2014.07 - 2014.11
  - 实现已有稀疏实现方法，进行对比实验
  - 研究稀疏学习方法在语音识别中的应用并实现
- 2014.12 - 2015.01
  - 探索语音识别的层次化网络结构
- 2015.02 - 2015.05
  - 确定硕士毕业论文提纲，完成实验，撰写毕业论文

# 参考文献

- G. Sivaram, H. Hermansky, "Multilayer perceptron with sparse hidden outputs for phoneme recognition", ICASSP'11, pp.5336 – 5339, 2011.
- D. Yu, F. Seide, G. Li, and L. Deng, "Exploiting sparseness in deep neural networks for large vocabulary speech recognition", In ICASSP'12, pages 5336 – 5339. IEEE, 2012
- Honglak Lee, "Unsupervised Feature Learning via sparse hierarchical representations", PhD Thesis, Stanford University, 2010.
- Ba, Jimmy, and Brendan Frey. "Adaptive dropout for training deep neural networks." Advances in Neural Information Processing Systems. 2013.
- Srivastava, Nitish. Improving neural networks with dropout. Diss. University of Toronto, 2013.
- Marc' Aurelio Ranzato, Y., Lan Boureau, and Yann LeCun. "Sparse feature learning for deep belief networks." Advances in neural information processing systems 20 (2007): 1185-1192.

# 参考文献

- Dahl, George E., et al. "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition." *Audio, Speech, and Language Processing, IEEE Transactions on* 20.1 (2012): 30-42.
- Hinton, Geoffrey, et al. "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups." *Signal Processing Magazine, IEEE* 29.6 (2012): 82-97.
- Jaitly, Navdeep, et al. "Application of Pretrained Deep Neural Networks to Large Vocabulary Speech Recognition." *INTERSPEECH*. 2012.
- Sivaram, Garimella SVS, et al. "Sparse coding for speech recognition." *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010.
- Smit, W. J., and E. Barnard. "Continuous speech recognition with sparse coding." *Computer Speech & Language* 23.2 (2009): 200-219.
- W. J. Smit, E. Barnard, "Continuous speech recognition with sparse coding", *Computer Speech & Language*, vol.23, no.2, pp.200-219, 2009.

# 谢谢！

- Q&A