学校编码：10384

学号：23120101152950

厦门大学

硕士学位论文

基于稀疏表示的说话人确认方法研究

Research of Speaker Verification System Based On Sparse Representation

王薇

指导教师：李琳

专业名称：电路与系统

答辩日期：2013年5月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下，独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为（　　）课题（组）的研究成果，获得（　　）课题（组）经费或实验室的资助，在（　　）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）：

年　月　日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文(包括纸质版和电子版)，允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于:

(　　)1.经厦门大学保密委员会审查核定的保密学位论文，于　年　月　日解密，解密后适用上述授权。

(　　)2.不保密，适用上述授权。

(请在以上相应括号内打“✓”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。)

声明人(签名):

年　　月　　日

摘 要

说话人识别作为现代生物信息识别中的一项重要技术，依据语音信号确认说话人身份。从1999年起，历年NIST测评结果显示，GMM-UBM识别框架使用统一背景模型自适应出目标说话人模型，能更好地表征说话人个性特征。由于GMM建模只是对目标说话人一类数据进行的，直接采用GMM似然度得分进行分类具有计算量大、区分能力不佳等不足之处；将GMM均值超向量作为SVM分类器的输入，采用非线性核函数进行二分类，一定程度上提高了说话人识别性能，但是数据的不平衡和两类数据的混叠问题对分类效果影响较大。稀疏表示理论指出可压缩信号能够在某个空间中由最能反映信号特征且数量最少的原子线性表示，表征同类信号的基原子分布密集，对不同类信号的表征具有很强的区分性，可实现信号的分类。本文深入研究语音信号的稀疏表示方法，实现了基于稀疏表示的说话人确认系统。本文主要工作及取得的成果如下：

(1)、系统分析了稀疏表示理论，对语音信号的稀疏表示方法进行了研究。鉴于语音信号的稀疏性和可重构性，对语音信号进行稀疏表示是可行的，并且稀疏表示方法采用分布密集的基原子表征同类信号，对不同类信号的表征具有很强的区分性。因此，利用稀疏表示对语音信号进行分类，能够取得很好的效果。

(2)、提出了基于稀疏表示的说话人确认系统GMM-SRC，在GMM-UBM基线系统的基础上，使用目标说话人和背景说话人的GMM均值超向量构建字典，通过 范数最优化方法求解稀疏表示系数，再利用重构误差进行识别。在NIST06年语音数据库上进行说话人确认实验，结果显示，GMM-SRC系统的等错误率比基线系统降低了1.27%。进一步提出了GMM-SRC和GMM-UBM的说话人确认融合系统，在分数域将两种系统进行融合，相比于基线系统，等错误率降低了2.35%。

(3)、针对GMM-SRC说话人确认系统，提出了一种改进的字典学习方法，在字典的更新过程中加入分类信息，采用D-KSVD算法训练字典，使生成的字典同时具有稀疏表达和分类能力。进一步，实现了基于D-KSVD的说话人确认系统，其获得的等错误率相比于基线系统降低了3.32%。

关键词：说话人确认；稀疏表示理论；字典学习；D-KSVD

Abstract

Speaker recognition as an important technology in modern biological information recognition area, it can confirm the identity of speaker based on the speech signal. Since 1999, NIST speaker recognition evaluation results show that, GMM-UBM recognition framework which gets target speaker GMM model from universial background model (UBM) adaptively can better characterize the speaker's personality. However, GMM is just trained from the target speaker training data which is only one class, also, directly using GMM likelihood scoring to classify has shortages of large computation and poor classification ability. SVM classifier using GMM mean super vector as input, through non-linear kernel function carry out binary classification, to some extent, it improves the performance of speaker recognition. But SVM has much difficulty in handling the unbalance and great deal of data. Sparse representation theory proves that compressible signal could be seen as a linear combination of base elements from the same class which most of the coefficients are zero, can realize signal classification. In this article, we delved the sparse representation methods of speech signal and proposed a speaker verification system based on sparse representation. The main research contents and results are as follows:

(1).Analysed the sparse representation theory systematically and studied the sparse representation method of speech signal. Due to speech signal's sparse and reconfigurable ability, it is feasible to do sparse representation of speech signal. In addition, sparse representation theory proves that compressible signal could be seen as a linear combination of base elements from the same class. Therefore, using sparse representation to classify speech signal can achieve well results.

(2).This thesis proposed a speaker verification system based on sparse representation GMM-SRC. On the basis of the GMM - UBM baseline system, we

built dictionary by the target and background speakers GMM mean supervector, then computed sparse coefficients by minimization, got the results by reconstruction error at last. Experiments on NIST06 showed that comparing with GMM-UBM baseline system, 1.27% relative reduction was achieved in EER. Further, we proposed a fusion system of GMM-SRC and GMM-UBM, comparing with GMM-UBM baseline system, 2.35% relative reduction was achieved in EER. (3). For GMM-SRC speaker verification system, this thesis proposed a new dictionary learning method. Adding classified information in the dictionary updating process, and then used the D-KSVD algorithm to train the dictionary. Allowing the performance of a linear classifier and the representational power of the dictionary being considered at the same time. At last, we proposed a speaker verification system based on D-KSVD. Comparing with GMM-UBM baseline system, 3.32% relative reduction was achieved in EER.

Keywords: speaker verificationsparse representationdictionary learningD-KSVD

参考资料

[1] Kersta L G. Voiceprint identification[J]. Police LQ. 1973, 3: 5.
[2] Atal B S. Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification[J]. the Journal of the Acoustical Society of America. 1974, 55: 1304.
[3] Doddington G R. Speaker recognition—Identifying people by their voices[J]. Proceedings of the IEEE. 1985, 73(11): 1651-1664.
[4] Higgins A, Wohlford R. A new method of text-independent speaker recognition[C]. IEEE, 1986.
[5] Soong F K, Rosenberg A E. On the use of instantaneous and transitional spectral information in speaker recognition[J]. Acoustics, Speech and Signal Processing, IEEE Transactions on. 1988, 36(6): 871-879.
[6] Deller J R, Proakis J G, Hansen J H. Discrete-time processing of speech signals[M]. Ieee New York, NY, USA, 2000.
[7] Oglesby J, Mason J S. Speaker recognition with a neural classifier[C]. IET, 1989.
[8] Rabiner L R. A tutorial on hidden Markov models and selected applications in speech recognition[J]. Proceedings of the IEEE. 1989, 77(2): 257-286.
[9] Juang B, Rabiner L R. Hidden Markov models for speech recognition[J]. Technometrics. 1991, 33(3): 251-272.
[10] Reynolds D A, Rose R C. Robust text-independent speaker identification using Gaussian mixture speaker models[J]. Speech and Audio Processing, IEEE Transactions on. 1995, 3(1): 72-83.
[11] Reynolds D A. Speaker identification and verification using Gaussian mixture speaker models[J]. Speech communication. 1995, 17(1): 91-108.
[12] Schmidt M, Gish H. Speaker identification via support vector classifiers[C]. IEEE, 1996.
[13] Wan V, Campbell W M. Support vector machines for speaker verification and identification[C]. IEEE, 2000.
[14] Wan V, Renals S. Evaluation of kernel methods for speaker verification and identification[C]. IEEE, 2002.
[15] Wan V, Renals S. Speaker verification using sequence discriminant support vector machines[J]. Speech and Audio Processing, IEEE Transactions on. 2005, 13(2): 203-210.
[16] Bengio S, Mari é thoz J. Learning the decision function for speaker verification[C]. IEEE, 2001.
[17] Fine S, Navratil J, Gopinath R A. A hybrid GMM/SVM approach to speaker identification[C]. IEEE, 2001.
[18] Fine S, Navr á til J, Gopinath R A. Enhancing GMM scores using SVM " hints ", "[C]. Citeseer, 2001.
[19] Liu M, Dai B, Xie Y, et al. Improved GMM-UBM/SVM for speaker verification[C]. IEEE, 2006.
[20] Reynolds D A, Quatieri T F, Dunn R B. Speaker verification using adapted Gaussian mixture models[J]. Digital signal processing. 2000, 10(1): 19-41.
[21] Kenny P, Boulianne G, Ouellet P, et al. Joint factor analysis versus eigenchannels in speaker recognition[J]. Audio, Speech, and Language Processing, IEEE Transactions on. 2007, 15(4): 1435-1447.
[22] Garcia-Romero D, Espy-Wilson C Y. Analysis of i-vector length normalization in speaker recognition systems[C]. 2011.
[23] Campbell W M, Sturim D E, Reynolds D A, et al. SVM based speaker verification using a GMM supervector kernel and NAP variability compensation[C]. IEEE, 2006.
[24] Martin A F, Przybocki M A. The NIST speaker recognition evaluations: 1996-2001[C]. 2001.
[25] 韩纪庆,张磊,郑铁然. 语音信号处理[M]. 清华大学出版社有限公司, 2004.
[26] Martin A, Przybocki M. The NIST 1999 speaker recognition evaluation—an overview[J]. Digital signal processing. 2000, 10(1): 1-18.
[27] Naseem I, Togneri R, Bennamoun M. Sparse representation for speaker identification[C]. IEEE, 2010.
[28] Donoho D L. Compressed sensing[J]. Information Theory, IEEE Transactions on. 2006, 52(4): 1289-1306.
[29] Mallat S G, Zhang Z. Matching pursuits with time-frequency dictionaries[J]. Signal Processing, IEEE Transactions on. 1993, 41(12): 3397-3415.

[30] Chen S S, Donoho D L, Saunders M A. Atomic decomposition by basis pursuit[J]. SIAM review. 2001, 43(1): 129-159.
[31] Peyre G. Best basis compressed sensing[J]. Signal Processing, IEEE Transactions on. 2010, 58(5): 2613-2622.
[32] Pati Y C, Rezaiifar R, Krishnaprasad P S. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition[C]. IEEE, 1993.
[33] Abdallah S A, Plumbley M D. Polyphonic music transcription by non-negative sparse coding of power spectra[C]. 2004.
[34] 余丰,吴尘. 基于压缩感知的稀疏线性预测语音编码[J]. 信息化研究. 2011, 2: 17.
[35] Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries[J]. Image Processing, IEEE Transactions on. 2006, 15(12): 3736-3745.
[36] Jafari M G, Plumbley M D. Fast dictionary learning for sparse representations of speech signals[J]. Selected Topics in Signal Processing, IEEE Journal of. 2011, 5(5): 1025-1031.
[37] Zhao N, Xu X, Yang Y. Sparse Representations for Speech Enhancement[J]. Chinese Journal of Electronics. 2011, 19: 21.
[38] Zibulevsky M, Pearlmutter B A. Blind source separation by sparse decomposition in a signal dictionary[J]. Neural computation. 2001, 13(4): 863-882.
[39] Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 2009, 31(2): 210-227.
[40] Kua J M K, Ambikairajah E, Epps J, et al. Speaker verification using sparse representation classification[C]. IEEE, 2011.
[41] Haris B C, Sinha R. Speaker verification using sparse representation over KSVD learned dictionary[C]. IEEE, 2012.
[42] Pati D, Prasanna S. Speaker information from subband energies of linear prediction residual[C]. IEEE, 2010.
[43] Hieu N T. Speaker diarization in meetings domain[J]. School of Computer Engineering, Nanyang Technological University, Proposal for admission to the Degree of Doctor of Philosophy Sept. 2009.
[44] Cheng Y, Leung H C. Speaker verification using fundamental frequency[C]. 1998.
[45] 姚志强,吴礼福,戴蓓倩,等. 用于文本无关的话者识别的超音段信息提取[J]. 数据采集与处理. 2005, 20(4): 376-380.
[46] Itakura F. Minimum prediction residual principle applied to speech recognition[J]. Acoustics, Speech and Signal Processing, IEEE Transactions on. 1975, 23(1): 67-72.
[47] Selvi T, Pragatheeswaran J. Efficient speech enhancement technique by exploiting the harmonic structure of voiced segments[C]. IEEE, 2011.
[48] Furui S. Cepstral analysis technique for automatic speaker verification[J]. Acoustics, Speech and Signal Processing, IEEE Transactions on. 1981, 29(2): 254-272.
[49] Viikki O, Laurila K. Cepstral domain segmental feature vector normalization for noise robust speech recognition[J]. Speech Communication. 1998, 25(1): 133-147.
[50] Hsu C, Lee L. Higher order cepstral moment normalization (HOCMN) for robust speech recognition[C]. IEEE, 2004.
[51] Pelecanos J, Sridharan S. Feature warping for robust speaker verification[J]. 2001.
[52] Hermansky H, Morgan N. RASTA processing of speech[J]. Speech and Audio Processing, IEEE Transactions on. 1994, 2(4): 578-589.
[53] Reynolds D A. Channel robust speaker verification via feature mapping[C]. IEEE, 2003.
[54] 刘明辉. 基于 GMM 和 SVM 的文本无关的说话人确认方法研究[D]. 合肥: 中国科学技术大学, 2007.
[55] Li K, Porter J E. Normalizations and selection of speech segments for speaker recognition scoring[C]. IEEE, 1988.

[56] Auckenthaler R, Carey M, Lloyd-Thomas H. Score normalization for text-independent speaker verification systems[J]. Digital Signal Processing. 2000, 10(1): 42-54.
[57] Reynolds D A. The effects of handset variability on speaker recognition performance: Experiments on the switchboard corpus[C]. IEEE, 1996.
[58] Boominathan V, Sri Rama Murty K. Speaker recognition via sparse representations using orthogonal matching pursuit[C]. IEEE, 2012.
[59] 黄韬,刘胜辉,谭艳娜. 基于 k-means 聚类算法的研究[J]. 计算机技术与发展. 2011, 21(7): 54-57.
[60] Tzortzis G F, Likas A C. The global kernel k-means algorithm for clustering in feature space[J]. IEEE transactions on neural networks. 2009, 20(7): 1181.
[61] Pham D, Venkatesh S. Joint learning and dictionary construction for pattern recognition[C]. IEEE, 2008.
[62] Teunen R, Shahshahani B, Heck L. A model-based transformational approach to robust speaker recognition[C]. 2000.
[63] Kenny P, Boulianne G, Ouellet P, et al. Improvements in factor analysis based speaker verification[C]. IEEE, 2006.
[64] Zhang Q, Li B. Discriminative k-svd for dictionary learning in face recognition[C]. IEEE, 2010