



高木 一幸
Kazuyuki TAKAGI

研究课题 采用非负矩阵分解的语言识别技术、用于语音输入·控制的语音识别技术

关键词

语言识别, 音素排列信息, 非负矩阵分解, 语音识别, 耐噪音性, 多频带·多信噪比·多通道语音识别, 模型识别, 机械学习

所属专业	研究生院信息理工学研究科 综合信息学专业
研究成员	高木 一幸 助教
所属学会	电子信息通信学会, 日本声学会, 信息处理学会, 人工智能学会, 日本声学会, IEEE, ISCA
研究设备	语音语言媒体信息处理系统

研究概要

语音识别技术的发展与普及

该研究室从事基于机械的语音语言处理及识别技术相关的研究。所谓语音识别是将人发出的声音转化成文字或者判断说话内容的一门技术。

在这几年里, 语音识别技术越来越多地融入到我们的生活中。比如搜索引擎 GOOGLE 就装载了语音识别功能, 智能手机等设备也能用语音输入。并且市场也有销售用语音控制室内空调、LED 照明、电视等家电产品的语音遥控器。语音已经不仅被用于人与人之间的交流, 还被应用到人与机械的交流中。

语音识别

该研究室特别着重研究的是语音识别。这项研究是用计算机来自动识别未知的说话者讲的是哪一国的语言。

美国等多民族国家对这种语言识别技术的应用需求非常高。比如向消防局及警察局拨打紧急电话时, 会存在平时讲英语的人由于惊慌而在无意识中使用母语的情况。为此美国的这些机构需要设置会讲不同语言的工作人员, 但是在电话中很难判断报警人员讲的哪一门语言, 判断一旦不及时就有可能导致重大问题的发生。这时如果能够利用语言识别技术, 就能将电话迅速准确地转给负责人。

像这样对应用自动语言识别技术服务的需求非常高。在不久的将来, 不仅是亚洲地区, 全世界都会需要这种技术。

由于语言的差异表现为音素类别、单词的音素排列模式、韵律模式、语汇等, 所以可利用与这些差异相关的语音信息来进行判断。比如日语的情况下, 它具有按元音、辅音、元音、辅音、元音、辅音这种顺序排列的特点, 总之是不可能出现辅音、辅音、辅音、元音这种排列。机械收录这些音素排列信息进行统计, 将这些语言的特征转化成模型。将输入的语音与这种模型进行比较来判明是哪门语言。

优势

应用非负矩阵分解的语言识别技术

该研究室开发的语音识别技术的优势在于提取音素类别、音素排列的特征量时应用了非负矩阵分解 (Nonnegative Matrix Factorization: NMF)。NMF 是进行图像解析及从乐器声中自动采样时使用的一种方法, 将这种方法应用到语音中时, 如果事先学习音素的矢量特性, 就能按时序知道会出现什么样的音素。用这种方法不断地提取声音就能解析语音的特点。

这里是利用被用作语言模型的 N-gram (n 个要素的并列类别和频率的统计) 将提取的音素和显现

的信息转化模型, 最后将模型的参数作为特征矢量, 用支持向量机等学习机械来进行语言识别。这种语言分析所采用的 NMF 技术是该研究室原创的方法和技术, 几乎还没有其他研究室展开这样的研究。

用 NMF 实现 98.6% 的语言识别率

开始使用 NMF 的动机是基于 2008~2010 年与声学 and 语言学研究人员共同研究多语言处理时所积累的经验。根据共同研究所获取的各种知识和经验, 他们想到了能够活用这种方法。

使用这种方法来识别日语与英语时, 结果表明在选取最佳的音素频谱模板的数量与 N-gram 的次数时, 实现了 98.6% 的识别率。除此之外, 该研究室还已经拥有了约 20% 的语言数据库, 所以今后还会将这种方法逐步应用到各语言中。

耐环境噪音的语音识别

该研究室还从事耐环境噪音的语音识别技术的研究。在实际应用语音识别时周围会存在各种噪音, 由于这些噪音的存在, 需要保持语音识别的精度。

该研究室采用电气通信大学尾関和彦研究室研究的多频带、多信噪比、多通道语音识别法, 继续研究即使在较多噪音的环境下也能以较高的识别率识别单词语音的方法。

多频带、多信噪比、多通道语音识别法是识别每个频带内的语音。比如噪音小的情况下很难进行低音语音识别, 但却能准确地进行高音语音识别。每个频带的语音识别采用了声学模型, 该模型是按各种信噪比 (SNR) 将系统所学习的语音和杂音混杂在一起而作成的。噪音实际上多种多样且时常变化。根据每个频带的情况酌情选择与输入语音最匹配的声学模型进行语音识别, 通过将选择的声学模型进行组合, 语音识别的整体概率就会很高。

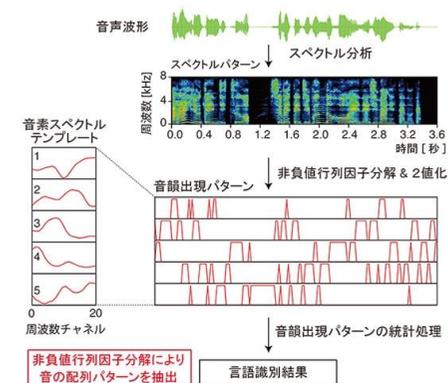
未来展望

希望缔造出翻译机器人的头脑

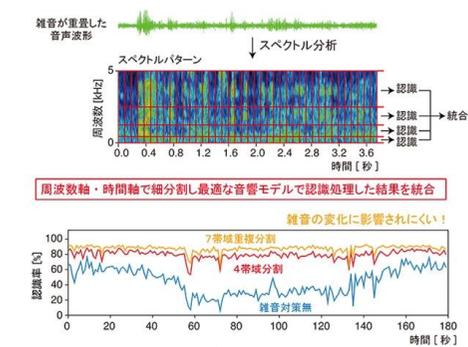
长期的研究结果表明为了进一步提高单词识别的精度, 需要进一步改良频带的分割方法, 在制作声学模型时在所用的模型噪音的特性上想办法即可, 所以该研究室希望在这个方向上加深研究。语言识别的研究于 2011 年取得了较好的成果。他们的语言识别技术因为使用了别无他例的独特方法, 因此期待着采用世界标准的评价数据能将精度提高到什么程度。并且将来他们还希望缔造出出现在电影里的翻译机器人的头脑。其中特别重视多语言识别, 他们的梦想和目标是不仅要判断出所讲的是哪国语言, 还能识别各国的地方语言及方言。



语音语言媒体信息处理系统的 UNIX 服务器



采用非负矩阵分解 (NMF) 来提取特征的语音识别



雑音の変化による単語認識率の変化
雑音= 路線バス車内、話者= 男性、単語数= 281、SNR= 0 dB
多频带・多 SNR・多通道语音识别方法耐噪音变化