技术调研：深度学习On智能家居（手机应用）

朱文韬 1600012785

一、应用现状

智能家居的概念源于人们对舒适便捷生活方式孜孜不倦的追求，融合了微电子、网络通信、自动控制、人工智能等多学科技术，具有相当高的综合性。近年来随着“物联网”概念的兴起，智能家居越发得到产业界的重视，融合深度学习的技术应用也得到了一定的发展。

就深度学习在智能家居的应用而言，从产品上主要可以分为平台类和工具类应用（前者如HomeKit、Amazon Echo、Google Home，后者如Ecobee智能温控器、Nest Cam安防摄像头等）；从技术上可以分为人机交互、用户行为识别、用户行为预测几大类。

手机常常被设计为智能家居的用户直接输入端，因此需要具备良好的交互特性。虽然Amazon、Google正在尝试以智能音箱为主要交互媒介切入智能家居市场，但手机的高使用率、便携性使其在智能家居的控制端（尤其是在远程控制场景下）仍具有不可替代的地位。同时，手机也可以作为智能家居的一个控制中枢而存在，如利用手机所具有的一定的计算能力统计、分析家庭中各电器、传感器的数据，从而实现对用户行为的整体识别乃至预测。在这些问题中，深度学习都具有难以替代的价值。

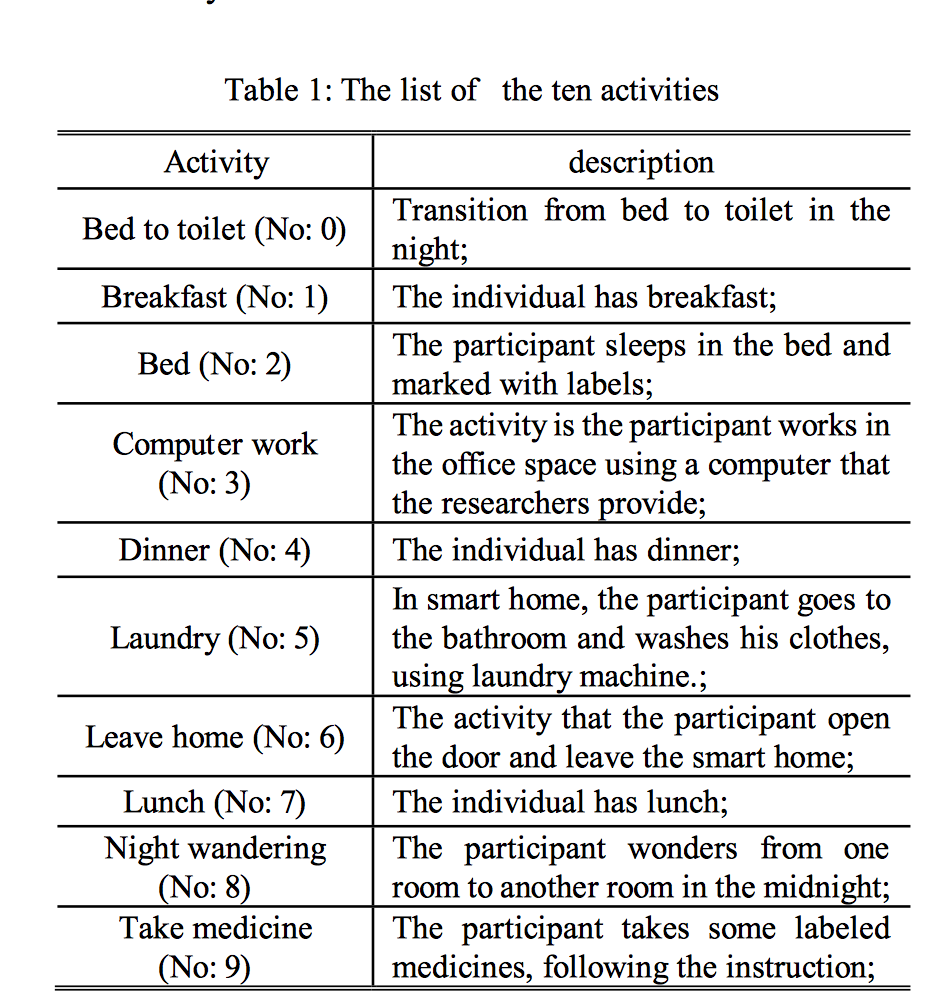
1.1人机交互

鉴于手势控制、眼动控制往往依赖更加苛刻的外部硬件（如搭载深度红外相机的Kinect、眼动仪EyeTribe等）因而尚没有在智能手机中得到广泛应用，语音识别成为智能家居控制中人机交互的主要实现方式。

在将音频送入神经网络之前首先经过预处理：降噪后的音频通过采样、傅立叶变换筛选频带后输入神经网络，神经网络将为每一个切分的小音频片段对应最有可能的语素，得到一系列最有可能的翻译结果。进一步，根据训练数据对这些基于发音的预测进行打分。初步输出的转录可能与音频的真实语义相差很远，但经过处理后能得到最接近的转录文本。

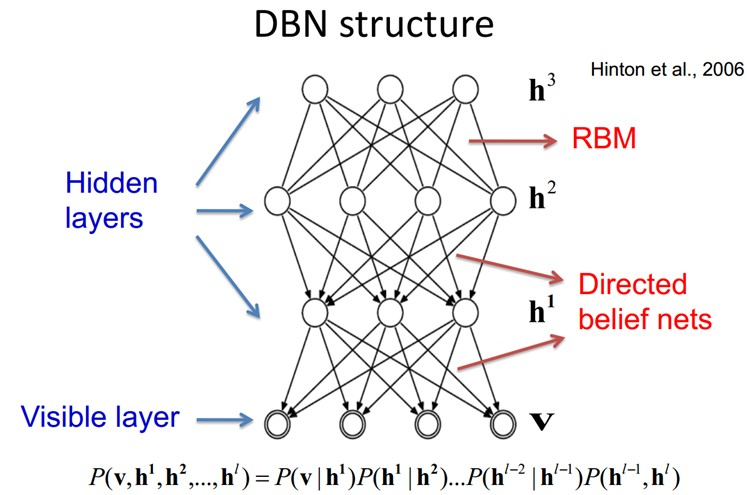
1.2 用户行为识别

理想的智能家居的中央控制端还应具有用户行为识别的功能，即通过分散在家庭中的传感器和智能家电获取用户的活动信息，并从中抽象出用户处于何种行为状态，进而调整家庭中的联网设备进入合理的模式。这种抽象推理的过程通常借助深度学习技术来实现。



对于用户行为的识别问题可以采用Hinton提出的Deep Belief Networks (DBNs)有监督学习方法，其中往往设置4个隐层，其中顶部的每一个节点代表了一种用户行为。

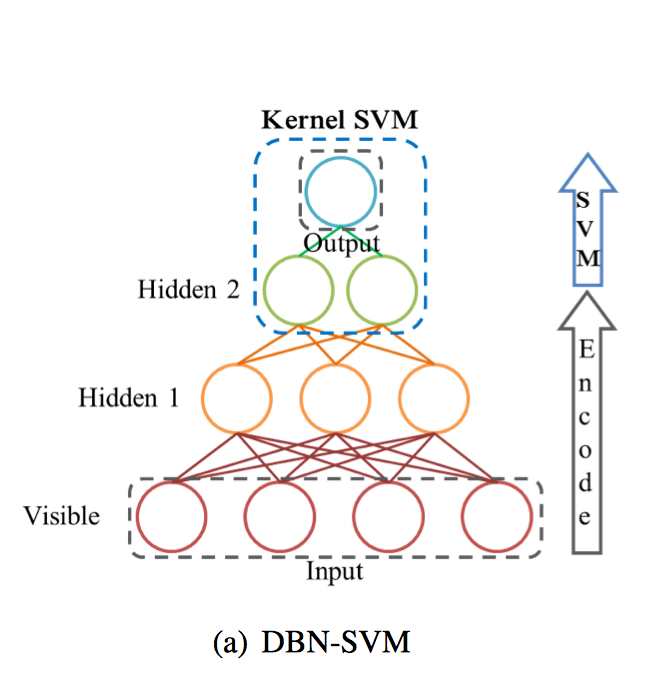
事实上，DBNs由多个限制玻尔兹曼机（Restricted Boltzmann Machines）层组成，一个典型的神经网络类型如图所示。



这些网络被“限制”为一个可视层和一个隐层，层间存在连接，但层内的单元间不存在连接。隐层单元被训练去捕捉在可视层表现出来的高阶数据的相关性。 RBM是一个简单的两层模型，每一层的节点之间没有连接，且所有的节点都是随机二值变量节点（只能取0或者1值）。随机选取两层中的一个状态为初始状态应用Gibbs采样。以在训练集上的正确识别率为目标函数训练这个网络，可以得到层间合适的系数。训练时根据已标注的用户活动数据（包含传感器、电器类型、状态及活跃时间）先后经过预实验、微调得到训练模型。和其他机器学习方法（如隐式马尔可夫模型、Naive贝叶斯分类器）相比，基于深度学习的用户行为分类方法具有最准确和稳定的表现。

1.3 用户行为预测

基于对用户行为的识别技术，预测用户的行为便成为了可能。一个普遍的预期是，用户的行为中蕴含着意图，而借助深度学习挖掘隐含意图可以预测用户未来的动向，从而提前安排电器的工作状态，可以实现节能和更好的用户体验。



一种典型的预测方法是使用Gibbs采样的限制玻尔兹曼机，在RBM上考虑一个最大似然估计问题（可以用梯度下降法迭代计算）。又有研究者提出将DBN-SVM（一种深度信念网络和支持向量机结合的方法），在MIT训练集上表现十分出色。

值得注意的是，现阶段对于用户行为预测的准确率还没有达到令人满意的程度，这既与训练集的覆盖程度有关，也与用户行为本身具有相对较高的复杂性、随机性有关。

二、应用前景

从人机交互的角度，智能家居的应用场景促进了近年来对远场语音信号处理的研究。远场语音信号处理是在利用深度学习的识别之前，在预处理模块尽可能地消除远距离拾音,解决噪声、混响、回声带来的影响，从而实现更加自由便捷的用户控制体验。远场语音识别技术难点可以分为3个部分，第一个是多通道同步采集硬件研发，第二个是前端麦克风阵列降噪算法，第三个是后端语音识别与前端信号处理算法的匹配。从目前的产品上看，麦克风阵列信号处理的方式在实际中用的比单纯的深度学习方法应用更多一些，Echo用了7个麦克风，叮咚用了8个，Google home也用了两个。大多数研究者认为麦克风阵列信号处理是有必要的，考虑到人除了大脑外还有双耳接受、处理声音，而目前深度神经网络还远远不能和人脑相比，因而更需要多个麦克风的辅助。这也是一定程度上各大科技公司除了手机之外选择使用基于智能音箱的语音交互切入智能家居市场的一个主要原因。

可以预期，在预处理降噪、对应到音素、翻译结果打分等语音识别的各个环节，不同的深度学习技术将得到应用、优化、整合。同时，在应用层面上，随着训练数据的增加，对多语言支持、根据上下文的自动更正将会获得更好的表现。事实上，在刚刚结束的Google I/O 2017大会中，谷歌宣布Google Home的语音识别错误率降到4.9%并能够支持同时识别6个人的语音。此外可以预期，随着更多智能硬件的进一步整合，手势识别、眼动控制等技术将更加成熟并被引入智能家居的日常控制，实现更加自然，同时不失精准的控制体验。

从用户行为识别的角度，Apple的Homekit智能家居生态系统自iOS10正式上线Automation自动化功能，用户可以设置针对特定的时间、地点、用户行为，触发中枢控制设备（iPhone/iPad/Apple TV）切换状态模式，并控制通过局域网连接的各智能设备。从苹果与富力地产合作，配置了诸多 Homekit 配件的精装修样板间来看，这个功能已经显著的优化了居住体验。然而，这只是用户行为识别的初级阶段，几乎没有涉及机器学习技术。可以预见的是，用户行为识别将从简单的自定义映射发展到能够主动分析行为、意图乃至情感的更高维的辅助系统，而只有达到这一步之后，才能够较好地分析“用户行为预测”的具体应用。在这一过程中将面临两个问题：

1、目前已经发布的智能家庭配件，大多还只是门锁、灯具等小型家电设备，替换频率更低、难度更大的大型家电还较罕有。而冰箱、卫浴、洗衣机等大型家电实行的标准、数据流的规范更难统一，产生的数据流量也远远更高，如何选取统一有效的分析模式将成为挑战。

2、针对智能家庭隐私保护的考虑，用户数据在云上的训练、计算带来的安全风险将不能被忽视。事实上，苹果为Homekit选择了被称为“端对端”的数据模式，开关门时间、净化器使用之类的用户数据，只会在 iPhone、iPad 设备与家居配件之间端对端通信。信息只会被存储于设备中，而不会被上传至苹果的任何云服务，以确保家庭环境的安全感和私密感，不会被智能化的家居配件所破坏。这就要求在进阶的用户行为识别中，深度学习系统可能更多地被部署在本地而非云端。

参考文献：

[1] Recognizing Human Activity in Smart Home Using Deep Learning algorithm. Proceedi ngs of the 33rd Chinese Control Conference July 28-30, 2014, Nanjing, China

[2] Human Behavior Prediction for Smart Homes Using Deep Learning.

Sungjoon Choi, Eunwoo Kim, Songhwai Oh.

[3] Smart Home: Cognitive Interactive People-Centric Internet of Things. Shuo Feng, Peym an Setoodeh, and Simon Haykin.

[4] https://www.apple.com/ios/home/

[5] https://events.google.com/io/schedule/?section=may-17