The Speech Recognition Virtual Kitchen: Launch Party

Andrew Plummer, Eric Riebling, Anuj Kamar, Florian Metze, Eric Fosler-Lussier, and Rebecca Bates

1Dept. of Computer Science and Engineering, The Ohio State University, Columbus, OH; U.S.A.
2Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA; U.S.A.
3Dept. of Integrated Engineering, Minnesota State University, Mankato, MN; U.S.A.

{plummer.321,fosler-lussier.1}@osu.edu, {erlk,fmetze,anujk1}@cs.cmu.edu, bates@mnsu.edu

Abstract

We present updates to the Speech Recognition Virtual Kitchen (SRVK) environment, a repository of pre-configured Virtual Machines (VMs) containing tools and experiments in the speech and language field. SRVK promotes community sharing of research techniques, fosters innovative experimentation, and provides solid reference systems as a tool for education, research, and evaluation. VMs provide a consistent environment for experimentation, without requiring tedious installation of many individual tools, a web-based community platform complements the VMs, allowing users to jointly explore, learn and collaborate using VMs. In this Show&Tell demo, we present the infrastructure to the speech community, along with several example VMs and a set of online error analysis tools. We solicit feedback from the community, in order to further guide development of the kitchen, which we hope to grow into a widely used community resource.

Index Terms: speech recognition, virtualization, educational tools, research infrastructure

1. Introduction

The depth and breadth of disciplines related to Automatic Speech Recognition (ASR) research and education has long reached a point where serious attention to community organization and infrastructure is of critical importance to its continued development and growth, and potential cross-disciplinary expansion. The following facets face significant challenges:

Basic ASR research and education: Speech recognizers incorporate knowledge from linguistics, phonetics, acoustics, signal processing, statistical modeling, graph theory, and artificial intelligence; expecting students to become experts in all of these areas, before attempting to work on speech recognition systems, is unrealistic.

Advanced ASR Research: Building and maintaining a state-of-the-art ASR system has moved beyond the ability of a single developer; it is difficult for all but the largest of university labs to build or maintain an end-to-end system, and adapt it to new conditions as required.

Cross-disciplinary ASR research and education: The challenges above pose a high bar for developing new research groups, making it difficult for institutions without active ASR researchers to integrate ASR projects into their educational curricula or field research projects which include ASR.

With the Speech Recognition Virtual Kitchen model, we extend the model of lab-internal knowledge transfer and infrastructure sharing to a community-wide effort through the use of Virtual Machines (Figure 1). We conceptualize virtual machines as a “kitchen” because they provide the infrastructure into which one can install “appliances” (e.g., speech recognition tool-kits), “recipes” (scripts for creating state-of-the art systems), and “ingredients” (language data). VMs share recipes in a ready-to-run fashion, together with data, log-files, results, etc. – a working environment that includes all tools and a reference or baseline, with links to other users that work on exactly the same task, anywhere in the world. Students and researchers alike can modify recipes step by step, observing the effect of changes. They can simply test a system on different data (e.g., different acoustics, speakers), or re-train a system in a different language, or with different data, observing how performance changes. A “kitchen” ASR system is therefore never a black box, but is extremely flexible, has good performance, and can easily be integrated into other, bigger projects, facilitating research for speech experts and non-experts alike.

2. The Speech Recognition Virtual Kitchen

The Speech Recognition Virtual Kitchen [1] provides the organization and infrastructure for meeting the numerous challenges facing the ASR research community in two key ways. The first is through the establishment of a set of VMs with associated repositories that facilitate the exchange of VMs among members of the ASR community and other interested parties. The second is through a web-based community platform which complements the repositories, and allows physically disconnected users to jointly explore VMs, learn from each other, and collaborate. Using open-source resources, such as Debian Linux derivatives as a platform, Kaldi [2] as a recognizer, and the TED-LIUM corpus [3] as example data, for example, allows us to create an infrastructure that can be freely shared.

The Kitchen repository contains ‘Kitchen’ VMs and software packages prepared by the community and/or the maintainers of the SRVK that are known to work, and for which doc-
A Virtual Worlds VM [1] with a basic speech recognition function, teaching needs, and other desiderata. We will also collect ideas for other scenarios in which this infrastructure will be useful, including fields that are data intensive (synthesis, dialog systems, NLP, computer vision, data mining). This may be mutually beneficial, as incubating ASR in other fields by providing an easy-to-use, non-trivial research environment will boost the relevance of speech and language technologies across disciplines. We are interested in accumulating systems for virtualization from within the ASR community, as well as those from adjacent fields.

5. Acknowledgments

The authors would like to thank Jesse Seguin, Malav Bhavsar, and Siddharth Kothari for their contributions to this work.

SRVK is supported by the United States’ National Science Foundation under grants CNS-1305365, 1305319, and 1305215.

6. References