

Adding intelligent help to mixed-initiative spoken dialogue systems

Genevieve Gorrell, Ian Lewin and Manny Rayner

Netdecisions Technology Centre

Westbrook Centre

Milton Road

Cambridge CB4 1YG

United Kingdom

{genevieve.gorrell, ian.lewin, manny.rayner}@netdecisions.com

1 Introduction

The demo illustrates an approach to providing help in a mixed initiative spoken dialogue system. Where grammar-based recognition fails, a robust approach is used to select a help message that aims to give the user some guidance, such that their next utterance is more likely to be recognised. We call this approach “targeted help”.

The base system is the On/Off House (OOH) system, which is implemented using the Nuance Toolkit platform, and offers English spoken language control, via telephone, of about 20 devices in a simulated home. Device types include both on/off and scalar. The dialogue manager is implemented in Visual C++ using the Nuance DialogueBuilder API. The mode of operation is primarily user-initiative. The grammar offers coverage of a fairly broad range of language, including commands (“Turn on the heater”, “Turn off the light in the bathroom”), several types of questions (“Is the heater switched on?”; “What is there in the kitchen?”; “Where is the washing machine?”; “Could you tell me which lights are on?”), universal quantification (“Switch off everything in the bathroom”), conjunction (“Are the hall and kitchen lights switched on?”; “Switch off the radio, TV and computer”), ellipsis (“Turn on the cooker”... “now the microwave”) and pronouns (“Switch off the stereo and the hi-fi”... “switch them on again”). The system has been tuned over four or five iterations of user testing, and performs well enough to have been successfully demonstrated in public on several occasions, including

at NAACL 2001, Pittsburgh.

Targeted help has been added to the system such that whenever an utterance is not recognised above a certain confidence threshold using the grammar-based system, instead of the standard error message, “Sorry, try again,” being played to the user, some further processing is done. First, the utterance is passed to a statistical language model appropriate to the domain for a second recognition. The result of this recognition contains information such as what words the recogniser recognised, what confidence scores it places on those words, what confidence score it places on the entire utterance, etc. This result is used to create a feature set. A decision tree classifier is then used to classify the feature set and return the class. This class maps directly to an error message, which is played to the user before returning to the main loop of the application. So for example, if the user says “I want to know where the television is,” the error message played might be “I didn’t quite catch that. To find out where a particular device is, you could try something like ”where is the radio”.”