A Prototype Reading Coach that Listens

Jack Mostow, Steven F. Roth, Alexander G. Hauptmann, and Mathew Kane

Critiqued by Art Munson DUE: February 2, 2004

1 Engineering for Non-Standard Speech

To handle false starts the authors extended the speech recognizer's lexicon to include truncated words. Similarly, they used a language model that expected repeated words (or repeated truncations) to model repetitions and that allowed any word within the same sentence as successors to model omissions. All of these "clever hacks" helped coerce the speech recognizer to handle disfluent speech.

I found these solutions particularly interesting. At a practical level, an engineering solution enabled a general purpose speech recognizer to meet the needs of a specific problem. At a higher level, loosening the constraints of the language model and increasing the range of expected input resulted in a more robust system that could better handle the reading habits of people (children, in this case). I think this approach might be applied to handle other areas of NLP where every day language use does not exactly match theoretical models. For example, one approach to understanding an ungrammatical utterance would be to measure how far the utterance is from known grammatical structures, and choosing the closest approximation. Some research could be done to compare different metrics for similarity / dissimilarity.

In NLP (and AI), using "fuzzy" rules and constraints seems to be advantageous rather than detrimental. The performance of decision trees is often helped by pruning the trained classifier. In contrast, other fields of computer science require precision. Any inexactitude in synchronization constructs would be disastrous. Fuzziness in security would also be unacceptable.

2 Metric for False Alarms Alarmingly Unsatisfactory

The authors measure Emily's ability to detect misread words using two metrics:

$$\begin{aligned} \text{detection rate} &= \text{(misread words detected) / (total words misread)} \\ \text{false alarm rate} &= \text{(number of false alarms) / (number of words read correctly)} \end{aligned}$$

They report a detection rate of .488 (40 of 82) and a false alarm rate of .0366 (187 of 5106). These numbers look good on the surface, especially given the authors' arguments that even partial detection of misread words can be a big help to coaching reading. While the formula for the detection rate makes sense, using the *number of words read correctly* as the denominator for the false alarm rate trivializes the frequency of the false alarms. Certainly it is the correct ratio for detemining the percentage of the time the system has false alarms. It does not, however, convey the fact that Emily issues more than twice as many false alarms as total misread words (187 / 82). In other words, 82% of Emily's corrections are false alarms. Without a doubt, Emily is much better in this area than the predecessor system, Evelyn. Nonetheless, more work needs to be done to reduce the proportion of false alarms to true misread words detected. As the authors admit, "To avoid swamping the reader with unnecessary interventions, we must reduce the false alarms."