Confidence Measures for Speech Recognition

Reza Sadraei

Confidence Measures

- In speech recognition, confidence measures (CM) are used to evaluate reliability of recognition results.
- Confidence measures can help ASR systems to migrate from laboratory to real world.

Overview

- The approaches for computing CM can be presented as three major categories:
  - CM as a combination of predictor features
  - CM as a posterior probability
  - CM as utterance verification

Utterance Verification

- Utterance verification (UV) is a procedure used to verified how reliable are the results.

- Usually utterance verification is considered as a statistical hypothesis testing problem.

Utterance Verification

- Posing the problem in this fashion:
  - Does the input speech (*X*) contain the keyword corresponding to the most likely keyword (*W*) as determined by the speech recognizer?

Utterance Verification

- For a typical pattern classifier, given an observation *X* as input, we always get a pattern class *W* as output.

- *X* could come from several sources:
  - *X* actually comes from the class *W*
  - *X* comes from other classes instead of *W*
  - *X* is an outlier

Neyman-Pearson Lemma

- According to Neyman-Pearson lemma, an optimal test is to evaluate a likelihood ratio between two hypothesis H_0 and H_1.
  - H_0: X is correctly recognized (Null Hypothesis)
  - H_1: X is wrongly recognized (Alternative Hypothesis)

$$\text{LRT} = \frac{p(X|H_0)}{p(X|H_1)} \underset{H_1}{\overset{H_0}{\gtrless}} \tau$$

  - Where τ is the decision threshold.

Difficulty

- Computing null hypothesis is straightforward but the alternative hypothesis is a composite one, so that it is always very difficult to model H_1.

First Approach

- The likelihood ratio can be written as:

$$LRT = \frac{P(X \mid H_0)}{P(X \mid H_1)} = \frac{L(X \mid W)}{\frac{1}{N-1} \sum_{\substack{W' \\ W \neq W'}} L(X \mid W')}$$

  – Where *L(X/W)* is the likelihood of the observation *X* given pattern class *W*.

First Approach

- The models that are used for computing alternative hypothesis are called competing models.
- Computing LRT as it is defined, is required the evaluation of the likelihood of speech segment X for each of the models in the model set.
- To reduce the computational complexity, we can consider smaller number of competing models.

Second Approach

- The competing set (or cohort set) for a given *W* is defined to be a fixed number (K) of pattern classes that are most confusable with *W*.

Pattern Class	Cohort Set				
ae	eh	ay	aw	ey	ih
aw	ae	ah	ao	ow	aa
ao	aa	ow	aw	w	ah
m	n	ng	l	w	uw
ch	jh	sh	t	s	k
s	z	f	sh	th	h#
sh	s	ch	z	jh	zh

Second Approach

- The likelihood ratio can be written as:

$$LRT = \frac{P(X \mid H_0)}{P(X \mid H_1)} = \frac{L(X \mid W)}{\frac{1}{K} \sum_{W' \in CohortSet} L(X \mid W')}$$

Third Approach: UV based on Rival Model

- For a typical pattern classifier, given an observation *X* as input, we always get a pattern class *W* as output.

- *X* could come from several sources:
  - *X* actually comes from the class *W*
  - *X* comes from other classes instead of *W*
  - *X* is an outlier

- If an observation *X* is classified as *W* but it actually does not belong to the class *W*, we simply call it as a *rival* of the class *W*.

Third Approach: UV based on Rival Model

- The set of all rivals of *W*:

$$S_r(W) = \{X \mid L(W \mid X) > L(W' \mid X), \forall W' \neq W, X \not\subset W, and L(W \mid X) > \xi\}$$

- The set of observations from *W*:

$$S_c(W) = \{X \mid L(W \mid X) > L(W' \mid X), \forall W' \neq W, X \subset W\}$$

Third Approach: UV based on Rival Model

- The capability of utterance verification depends on how well we can distinguish $S_c(W)$ from $S_r(W)$.
- Statistical hypothesis testing can still adopted as a tool to separate $S_c(W)$ from $S_r(W)$ statistically.
- The simplest way to model $S_c(W)$ and $S_r(W)$ is that we estimate two different models Λ_c and Λ_r for $S_c(W)$ and $S_r(W)$, respectively, based on all possible training data from each of the sets.

Third Approach: UV based on Rival Model

- Once Λ_c and Λ_r are given, utterance verification is operated as the following likelihood ratio:

$$LRT = \frac{P(X \mid H_0)}{P(X \mid H_1)} = \frac{\Pr(X \in S_c(W))}{\Pr(X \in S_r(W))} = \frac{p(X \mid \Lambda_c)}{p(X \mid \Lambda_r)}$$

Third Approach: UV based on Rival Model

- It is straightforward to define $S_c(W)$ and $S_r(W)$ for every isolated word W.

- But for continuous speech recognition, it is very hard to associate a definite part of data to the rival set, because numerous boundaries are possible.

Using UV in Search Procedure

- It is possible to use utterance verification to correct some some possible recognition errors made by recognizer during search.
- At every time instant t, likelihood ratio testing is conducted for current path, if its score is below some threshold, this path will be rejectted.
- A wrong path with high likelihood but low verification score probability can be rejected during search.

Using UV in Search Procedure

- Another advantage of the above method is that likelihood ratio based confidence measure is calculated and attached with every phone in all possible paths.

- These phone scores can be easily put together to get the confidence measures for word, phrase, or the whole sentence.

Representing Results