

WS'98 Weekly Update

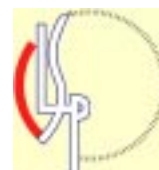
**ENHANCED SCORING METHOD FOR
ADAPTATION RESEARCH**

by:

Heather Collier
Sanjeev Khudanpur

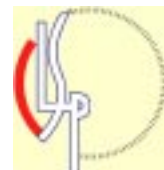
August 12, 1998

email: {heatherc, sanjeev}@clsp.jhu.edu



ENHANCED SCORING METHOD

- currently, the systems are scored based on an overall WER that is a combination of the percentage of insertions, deletions and substitutions generated over the entire test set
- the new method will look at this same data with respect to what data was seen during adaptation



ADAPTATION DATA

Unsupervised: Adapt on hypothesis

```
REF:  i saw * the dog BY THE river
HYP:  i saw A the dog BITE *** river
      C C  S C  C  I  D  C
```

"seen" words → correctly seen (C)
 incorrectly seen (I,S)
 missed (D)

Supervised: Adapt on reference

```
REF:  i saw the dog by the river
      C C  C  C  C  C  C
```

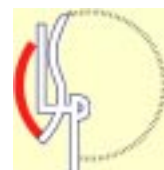
"seen" words → correctly seen (C)

TEST DATA

```
REF:  i * DUG a hole by THE river
HYP:  i A DOG a hole by *** river
      C I S  C C  C  D  C
```

For each word in test data REF and HYP:

- identify if seen in adaptation data
- classify into correct "seen" category
- adjust overall statistics (i.e. Ins, WER, etc.)
for that category



CENTER FOR LANGUAGE AND SPEECH PROCESSING

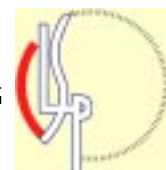
SMAP ADAPTATION RESULTS60s, Unsupervised Adaptation, $\tau=0.3$

BASELINE SYSTEM SUMMARY

all.rev.sys									
Adapt	HYP			REF			Total		
Presence	Ins	F.P.	Corr	Del	Sub	Corr	WER	Coverage	
Corr Seen	1.3	8.1	20.6	3.5	7.7	20.6	13.0	31.9	
Incorr Seen	1.0	6.8	9.9	2.0	4.6	9.9	9.9	16.5	
Missed	0.2	1.5	3.3	1.1	2.0	3.3	2.9	6.4	
Unseen	1.5	13.9	25.1	4.1	16.1	25.1	19.5	45.2	
Total	4.1	30.3	58.9	10.8	30.3	58.9	45.2	100.0	

ADAPTED SYSTEM SUMMARY

all.rev.sys									
Adapt	HYP			REF			Total		
Presence	Ins	F.P.	Corr	Del	Sub	Corr	WER	Coverage	
Corr Seen	1.4	8.2	21.5	3.1	7.2	21.5	12.8	31.9	
Incorr Seen	1.0	7.3	10.1	1.9	4.5	10.1	10.3	16.5	
Missed	0.3	1.6	3.3	1.1	2.0	3.3	3.0	6.4	
Unseen	1.5	13.3	24.8	3.7	16.7	24.8	18.5	45.2	
Total	4.2	30.4	59.7	9.9	30.4	59.7	44.5	100.0	



SMAP ADAPTATION RESULTS

60s, Unsupervised Adaptation, $\tau=0.3$

Facts:

Overall decrease in WER caused by:

- *increase* in WER for *incorrectly seen* words
- *decrease* in WER for *correctly seen* and *unseen* words

Hypotheses:

- increase in WER for incorrectly seen words due to increase in false positives (words substituted in HYP)
- decrease in WER for unseen words due to decrease in false positives

Why does this happen?

Baseline System:

REF: I saw X in X

HYP: I saw Y in Z

* *Y and Z are false positives*

Adaptation:

REF: When did the X appear

HYP: When did the Z appear

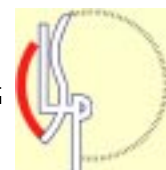
* *models for X adjusted to Z, an incorrectly seen word*

Adapted System:

REF: I saw X in X

HYP: I saw Z in Z

* *increase WER for incorrectly seen Z and decrease WER for unseen Y*



CENTER FOR LANGUAGE AND SPEECH PROCESSING

SMAP ADAPTATION RESULTS

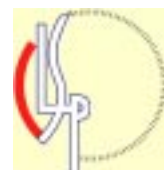
60s, Supervised Adaptation, $\tau=0.3$

BASELINE SYSTEM SUMMARY

all.rev.sys									
Adapt	HYP			REF			Total		
Presence	Ins	F.P.	Corr	Del	Sub	Corr	WER	Coverage	
Seen	2.4	15.2	33.0	6.7	14.8	23.0	24.3	54.5	
Unseen	1.7	15.1	25.9	4.1	15.6	25.9	20.9	45.5	
Total	4.1	30.3	58.9	10.8	30.3	58.9	45.2	100.0	

ADAPTED SYSTEM SUMMARY

all.rev.sys									
Adapt	HYP			REF			Total		
Presence	Ins	F.P.	Corr	Del	Sub	Corr	WER	Coverage	
Seen	2.4	14.5	35.5	5.8	13.1	35.5	22.7	54.5	
Unseen	1.5	14.1	26.2	3.8	15.5	26.2	19.4	45.5	
Total	3.9	28.6	61.8	9.6	28.6	61.8	42.2	100.0	



SMAP ADAPTATION RESULTS

60s, Supervised Adaptation, $\tau=0.3$

Facts:

Overall decrease in WER caused by:

- *decrease* in WER for *seen* words
- *decrease* in WER for *unseen* words
- *equivalent decrease* in WER for both seen words and for unseen words

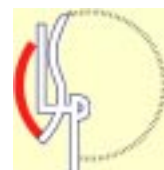
Why an equivalent decrease in WER?

- models are not actually adjusted at the word level:

words → phones → states

- adaptation is performed on the individual states
- even if you haven't seen a particular word in adaptation, you may have seen the separate states that make it up

cat → cattle



FURTHER WORK

- break the current code down to analyze similar results at the phone, state, or gaussian levels

