Porting: SwitchBoard to the VoiceMail Task

M.J.F. Gales, Y. Dong, D. Povey & P.C. Woodland
Cambridge University Engineering Department
{mjfg,yd211,dp10006,pcw}@eng.cam.ac.uk

INTRODUCTION

- Increasing need to rapidly generate systems for specific tasks. Either:
  - build a **generic** system which will work on a wide range of tasks;
  - **port** an existing "source" system to work on the target task.
- This paper examines various options for producing a system on limited task-specific training data.
- Extends previously published work (Cordoba et al ICASSP 2002). Examines:
  - **generic** systems, how to use task specific data?
  - **heteroscedastic LDA**: is a task specific projection useful?
  - **discriminative MAP**: a new porting/adaptation scheme, MMI-MAP (Povey et al ICASSP 2003) .

DATA SETS & EXPERIMENTAL SETUP

- **SwitchBoard** data used to train the **source** system (`swbd` data):
  - 265 hours of training data;
  - 4482 sides from Swb1 and 235 sides from CHE;
  - 3 hour test set (`dev01sub`) used to test genericity.
- **VoiceMail** used as the **target** task:
  - telephone messages addressed to IBM employees;
  - two training set releases (`vmtrain1` & `vmtrain2`) partitioned as:

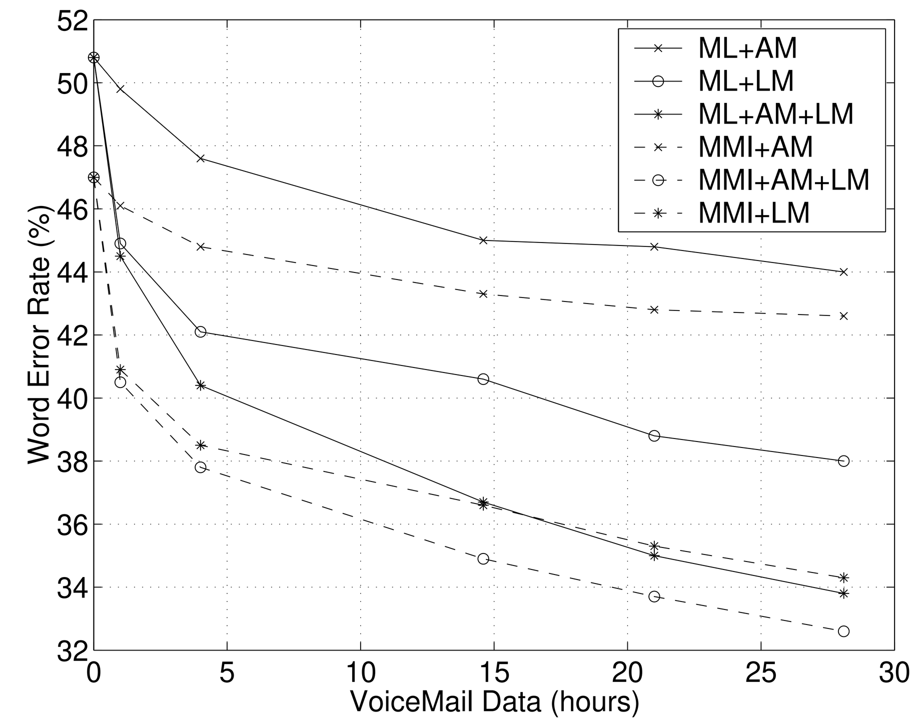
Subset	Amount	Comment
`1hr`	1.0hr	randomly selected from `vmtrain1`
`4hr`	4.0hrs	randomly selected from `vmtrain1`
`15hr`	14.6 hrs	`vmtrain1`
`20hr`	21.0 hrs	`vmtrain1` + 6.4hrs randomly selected from `vmtrain2`
`30hr`	28.1 hrs	`vmtrain1` + `vmtrain2`

  - test set (`vmtest`), 243 messages 94 minutes, comprising:
    1) `vmdevtest` from first VM data release (11mins);
    2) `vmtest2` from second VM data release (23mins);
    3) `vmtest3` heldout from second VM training data (60mins).
- **Wordlists**: three word lists used depending on LM adaptation:
  - no adaptation 27 thousand word vocabulary from SwitchBoard used;
  - using `1hr` `4hr` and `15hr` subsets for LM adaptation an additional 2 thousand words from the `15hr` subset added to initial list;
  - using `20hr` and `30hr` subsets for LM adaptation an additional 3.5 thousand words from the `30hr` subset added to initial list.
- **Language Model** data for baseline trigram LM (`Swbd-LM`):
  - 204 million words of Broadcast News data ;
  - 3 million words of `swbd` data transcriptions.

- **Source Acoustic Model System Setup**
  - MF-PLP to give 13 coefficients + delta+delta-delta parameters;
  - vocal tract length normalisation
  - mean and variance side-based normalisation;
  - gender independent models;
  - decision-tree state-clustered cross-word triphones;
  - 16 Gaussian components per state;
  - either ML or MMI parameter estimation used.

EXPERIMENTAL RESULTS

- **Baseline Porting Experiments**: (non-HLDA)
  - acoustic model (AM) porting using standard MAP, ML-MAP;
  - language model (LM) porting using interpolation (tuned on `vmtest2`);



  - greater gains from LM adaptation than AM adaptation;
  - gains from MMI initial model decreases as amount of VM data increases;
  - adapting AM and LM with `30hr` gave an error rate of 32.6%.
  - porting better than "pure" system for `30hr` subset (35.5%).
- **Generic System Generation**: (non-HLDA, no LM adaptation)
  - train on `swbd` and weighted `30hr` VM training data

	Test	`swbd+VM-30hr-×`				
		—	1	2	5	10
ML	`vmtest`	50.8	47.5	46.5	45.3	44.5
	`dev01sub`	38.1	37.8	37.8	38.1	38.4
MMI	`vmtest`	47.0	41.8	41.6	41.6	41.8
	`dev01sub`	34.8	34.9	35.1	36.0	36.4

  - ML training dramatically affected by data weighting;
  - weighting of 10 approximately matched `swbd` and VM hours;
  - MMI training gave better performance than ML-MAP adapted system (42.6%).
  - good performance on `dev01sub` compared to baseline ported system (39.5%).
- **Heteroscedastic LDA front-end**: (`30hr` subset, no LM adaptation)
  - use generic systems and HLDA to tune front-end to specific task;

HLDA Training	Acoustic Training	VM ×	WER (%) `vmtest`	WER (%) `dev01sub`
`swbd`	ML	—	49.2	37.0
	MMI	—	45.9	34.0
	ML	10	42.1	37.2
	MMI	1	39.5	34.1
`swbd+` `VM-30hr`	ML	10	42.4	37.5
	MMI	1	39.8	34.5
`VM-30hr`	ML	10	41.7	37.6
	MMI	1	40.2	36.0

  - `swbd` HLDA transform gave gains on all tasks;
  - no consistent gain from tuning front-end using HLDA.
- **Discriminative MAP**: (non-HLDA, no LM adaptation)

Adaptation	`1hr`	`4hr`	`15hr`	`20hr`	`30hr`
ML-MAP	46.1	44.8	43.3	42.8	42.6
MMI-MAP	46.1	44.0	41.4	41.1	40.5

  - MMI-MAP shows improvement over ML-MAP;
  - `30hr` performance better than generic MMI system (41.8%).
- **Combined Systems**: (`30hr` subset)
  - use interpolation to adapt LM as well as AM adaptation;

System	Style	Std	HLDA
ML (VM×10)	generic	34.1	32.6
MMI	generic	31.7	30.3
MMI	ML-MAP	32.6	31.4
MMI	MMI-MAP	30.8	29.7

  - best performance uses MMI-MAP with HLDA front-end.

CONCLUSIONS

- investigated forms of rapidly porting to a new task;
- generic systems with HLDA yield good performance;
- slight gains using discriminative adaptation (MMI-MAP);
- future work will concentrate on LM adaptation.