Boosting of contextual information in ASR for air-traffic call-sign recognition

Martin Kocour, Karel Veselý, Alexander Blatt, et al.

Brno University of Technology, Faculty of Information Technology
Božetěchova 1/2. 612 66 Brno - Královo Pole

ikocour@fit.vutbr.cz

September 2, 2021, INTERSPEECH

Call-sign boosting: The motivation

What we do?

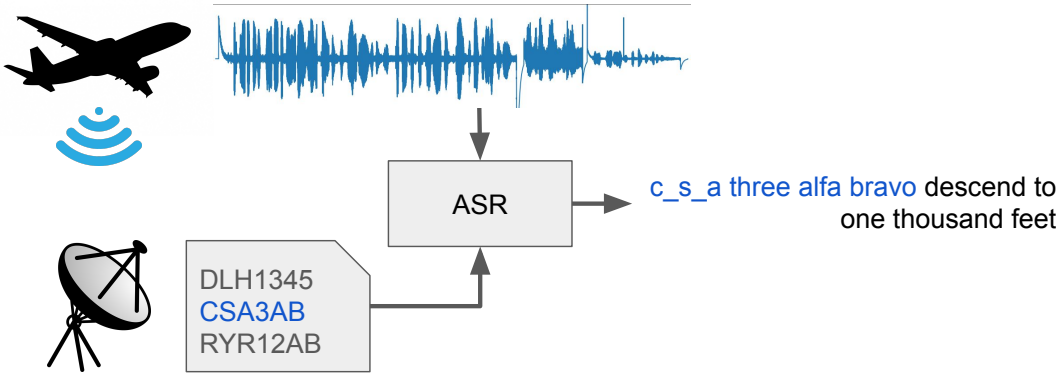
- **contextual adaptation** of ASR by means of WFST composition
- the context **changes rapidly** (from sentence to sentence)
- we focus on getting correct **call-signs** in ASR output

Call-sign = verbal identification of plane/flight in radio communication
Examples:

- c_s_a three alfa bravo, CSA3AB
- speedbird triple one, BAW111
- ryanair one two alfa bravo, RYR12AB

Call-sign = airline designator + letter/digit sequence

Call-sign boosting: The motivation



Goal: Improve call-sign recognition accuracy by leveraging the information from ADS-B messages of the airplane. Searched by location and time.

Call-sign boosting: The motivation

Challenging aspects of call-signs recognition:

- there are many airline designators: +/- 7700 (e.g. c_s_a, speedbird, lufthansa)
  some of them are similar (arrowhead, arrow jet, arrow, red arrows)
- "unofficial" variants of airline designators
  (lufthansa → hansa, speedbird → bird)
- shortening of callsigns is frequent
  (c_s_a one two three alfa bravo → c_s_a alfa bravo)
- verbalization variants: 0 → zero zero zero, triple zero, triple oh

Call-sign boosting: The motivation

Possible applications of call-sign recognition:

- tracking of communication (which airplane?)
- search in recordings (incident investigation)
- communication error detection (unknown call-sign was said)
- cockpit use (notifications for the pilot)

Lattice boosting

The contextual adaptation is done by Lattice boosting:

- Call-sign is first verbalized
  e.g. CSA3AB $\rightarrow$ c_s_a three alfa bravo
- Each such word sequence is boosted as a whole snippet of text
- Boosting is based on WFST composition

$$L' = L \circ B \tag{1}$$

- Each boosted sequence is represented as a specific path in the graph B, a path with score discounts that are added as offset to language model scores
- The graph B is replaced utterance to utterance
- The boosted lattice L' is then decoded to produce recognition output

Lattice boosting

- The graph consists of:
  - upper part – the boosted sequences with score discounts (-4/-6 per word)
  - lower part – the background model for not-matched paths
- Each boosted sequence is represented as a specific path in the graph
- Lower part is visited only if we cannot match with upper part (phi symbol #0)
- The word sequence must be present in original lattice in order to be boosted.

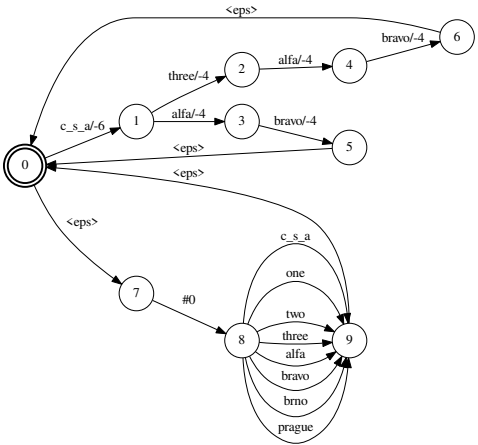


Figure: WFST of Lattice boosting graph *B*

Boosting of decoding network (HCLG)

We also boost HCLG graph for lattice generation (HCLG graph = recognition network)

- To increase chance that boosted sequence is present in the lattice
- Conceptually similar to lattice boosting, but topology of boosting graph is simpler

$$HCLG' = HCLG \circ B \qquad (2)$$

- Here we boost single words only (rare words), in our case airline designators (discount -3)
- Note `<eps>` transition back into the state 0
- affordable runtime of composition with HCLG thanks to trivial topology of boosting graph B

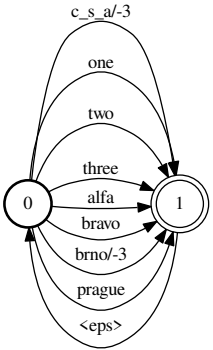


Figure: WFST graph *B* for HCLG boosting

Experiments

- Simulation of deploying ASR to new airports (unseen in training)
- Effect of call-sign boosting on call-sign recognition accuracy
  - liveatc_test_set2 data are noisy
  - malorca_vienna are clean and without pilots
  - call-sign recognition is done by neural network processing best hypothesis of ASR

Results

	Baseline		Lattice boost.		HCLG+Lat. boost.		Oracle
	CSA	WER	*CSA*	WER	*CSA*	WER	*CSA*
liveatc_test_set2	**53.5**	33.1	**75.6**	28.9	**80.6**	28.4	*90.0*
malorca_vienna	*84.4*	8.9	*86.5*	8.1	*88.1*	7.5	*90.5*
haawaii_bikf	-	30.6	-	29.4	-	28.9	-
haawaii_egll	-	20.8	-	19.3	-	18.8	-

Table: Call-Sign recognition Accuracy (CSA) % and Word Error Rate (WER) %

Conclusion

- Our cascade of HCLG and Lattice boosting helps to 'correct' the callsigns in the best ASR hypothesis.
- Our boosting method improves both the WER by -4.7% absolute and call-sign recognition accuracy by +27.1% absolute (liveatc_test_set2).

- Our cascade boosting technique is applicable also to other domains, if we know from the context the snippets of text that are likely to appear in ASR output and the likely rare words.
  - ☐ The context can change utterance to utterance.

http://atco2.org

http://haawaii.de

Thank You For Your Attention !