An Iterative Dual Pathway Structure for Speech-to-Text Transcription

Beatrice Liem
Haoqi Zhang
Yiling Chen

HCOMP 2011
audio transcription

“I have a dream that …”
audio transcription

1. Human transcription
   - Very accurate
   - High cost

2. Computer transcription
   - Less accurate
   - Lower cost

“I have a dream that …”
Crowdsourced transcription

CastingWords
TRANSCRIPTION SERVICES
Crowdsourced transcription

- Recruit Mturk workers to:
  - transcribe clips of audio
  - improve existing transcripts
  - grade transcripts

- Quality control
  - Grading the grader, multiple graders
  - It works! But, some wasted effort.
This work: remove checking process

- Iterative dual pathway structure incentivizes people to enter good transcriptions
- Embed iteration in a MapReduce framework
  (Dean and Ghemawat ’04; Kittur, Smus, and Kraut ’11)
MapReduce for audio transcription

“We hold these truths to be self-evident, that all men...”
Iterative tasks

(Little et al, ’10)

task $\rightarrow A_1 \rightarrow A_2 \rightarrow \ldots$
Iterative tasks

(Little et al, ’10)

\[
\text{task} \rightarrow A_1 \rightarrow A_2 \rightarrow \ldots
\]

\[
\text{task} \rightarrow \ldots \rightarrow A_2 \rightarrow \text{vote}(A_1, A_2) \rightarrow A_3
\]
Iterative dual pathway structure

A_1 \rightarrow A_2 \rightarrow \ldots

B_1 \rightarrow B_2 \rightarrow \ldots
Iterative dual pathway structure

score by similarity
to recent answers
on other path
Iterative dual pathway structure

- Combines iteration with output agreement
- Natural stopping condition by convergence
- Can implement in a game or as a Mturk task
Listen to the clip and write down what you hear.

You have been assigned to Team 2 for this clip. You will get points based on how well you match the transcript produced by the other team. You may use any transcripts generated by your team members (displayed below if applicable) to help you.

The real story behind the sheep is that Doctor Wilmouth created the prototype for bioindustrial design. He's the Henry Ford of the biotech century.
Study

- 147 Harvard undergrads

- Iterative version
  - 549 transcripts

- Parallel version (baseline)
  - 308 transcripts
Results

• Parallel: 93.6% word accuracy
• Iterative: 96.6% word accuracy
Example

Iteration 1: red, red, red! what should i do?

Iteration 2: red, red, red! Dear God, where should I go, what should i do?

Iteration 3: Fred, Fred, Fred! Dear God, where shall I go, what should i do?

Iteration 4: Rhett, Rhett, Rhett! Dear God, where shall I go, what shall I do? (Correct)
Feedback

- Most subjects claim to give high effort

- Iterative version was more enjoyable
“I have a dream that …”

Conclusion

Iterative tasks
+ Output agreement games
Thank you

Smell stew hq@eecs.harvard.edu

Smiles to hq@eecs.harvard.edu

Send mails to hq@eecs.harvard.edu