UB at the NTCIR-12 SpokenQuery&Doc-2: Spoken Content Retrieval Using Multiple ASR Hypotheses and Syllables

Jianqiang Wang

Department of Library and Information Studies

Graduate School of Education

University at Buffalo, the State University of New York

Background and Motivation

- MEI project at Summer 2000 JHU Workshop
  - Query-by-example: English news stories (AP and NYTimes) to retrieve Mandarin news broadcast (VOA)
  - Multi-scale paradigm: use of words, and subwords (Chinese characters and syllables)
- MALACH Project / CLEF-CLSDR track
  - Information access to Holocaust survivor testimonies: multilingual, spontaneous, oral history
- Online programs at UB/GSE/DLIS
  - MS in Information and Library Science, MS in School Librarianship: >90% courses are now online
  - Large volume of recorded lectures

Challenges and Techniques

- Challenges
  - ASR errors, topic boundaries, …
  - Information needs, users, user-system interfaces, …
- Techniques
  - (Most commonly) first converting speech to text and then applying text-based IR
  - Exploring rich features of ASR: words, subwords, time stamps, …

Our Areas of Interest

- Using multiple ASR hypotheses
  - Is it better than using only the "best" hypothesis?
- Using ASR syllables
  - Can it help when coupled with word-based retrieval?
- Comparing ASR engines (Julius and KALDI)
  - Do they result in different retrieval effectiveness?

Test Collection: Documents

98 conference lectures from SDPWS1-7

- Speech audio in wav format, divided into inter pausal units (IPUs)
- Manual transcriptions
- Reference automatic transcriptions, generated by two ASR engines (Julius and KALDI), at both word and syllable level
- Slide transition information
- Slide group segment information
- Slide to IPU alignment information

A “document” is defined as a slide group segment, which may contain one or more contiguous slides corresponding to a topic

Document Processing

- Creating document collections
  - Based on information of slide group segment and slide-to-IPU alignment
  - Multiple collections were created
    - Manual transcription
    - ASRs: word- or syllable-based, top-n hypotheses, Julius or KALDI
  - 2,259 documents per collection
- Segmenting Japanese text
  - Using MeCab Japanese Morphological Analyzer
- Converting into hexadecimal codes
  - For easy handling by the IR system
- Indexing each document collection
  - Multiple ASR hypotheses were treated as independent terms

Query Formulation

- Creating multiple query sets using
  - Manual transcription (verbose)
  - ASR texts: top-n words or syllables generated by the two ASR engines
- Segmenting
- Converting to hexadecimal codes

Average Document/Query Length

Term	Document	Query
Words of manual transcription	194	140
1-best ASR words by Julius	171	136
5-best ASR words by Julius	860	699
1-best ASR syllables by Julius	295	226
1-best ASR words by KALDI	174	180

IR System

Perl Search Engine (PSE): a Perl implementation of Okapi BM25 weighting

$$\sum_{w \in q}[\log \frac{(N - df(w) + 0.5)}{(df(w) + 0.5)}][\frac{(k_1 + 1) \times c(w,d)}{k_1((1-b) + b\frac{dl(d)}{avdl} + c(w,d))} \frac{(k_3 + 1) \times c(w,q)}{k_3 + c(w,q)}]$$

Where

k_1 = 1.2, b = 0.75, k_3 = 7

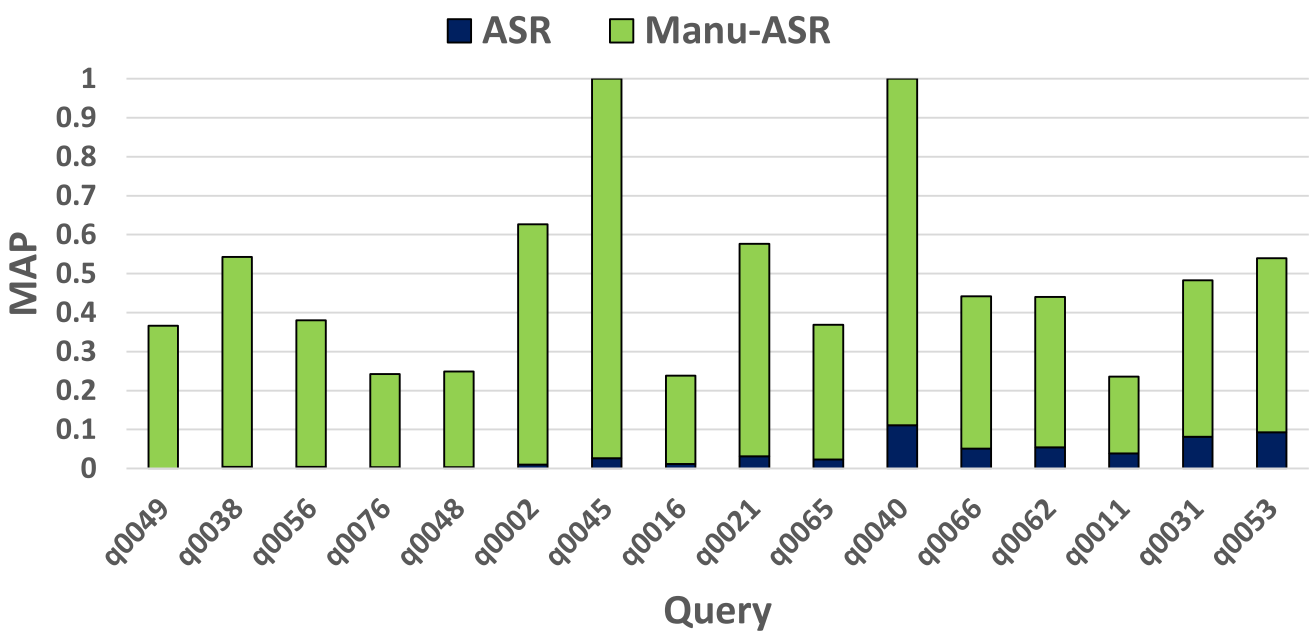
Official Runs and Results

Run id	Document term	Query term	MAP	Relative MAP
SQSCR-UB-SGS-TXT-1	Words of manual transcription	Words of manual transcription (verbose)	0.1953	Reference run
SQSCR-UB-SGS-TXT-2	1-best ASR words of Julius	1-best ASR words of Julius	0.1128	57.8%
SQSCR-UB-SGS-TXT-3	5-best ASR words of Julius	1-best ASR words of Julius	0.0994	50.9%
SQSCR-UB-SGS-TXT-4	1-best ASR words of Julius	5-best ASR words of Julius	0.1127	57.7%
SQSCR-UB-SGS-TXT-5	5-best ASR words of Julius	5-best ASR words of Julius	0.0966	49.5%
SQSCR-UB-SGS-TXT-6	1-best ASR syllables of Julius	1-best ASR syllables of Julius	0.0253	13.0%
SQSCR-UB-SGS-TXT-7	1-best ASR words of KALDI	1-best ASR words of KALDI	0.1946	99.7%

Comparison of Runs

- ASR vs manual transcription
  - Julius: significantly lower MAP regardless of index terms
  - KALDI: statistically indistinguishable MAP on 1-best words
- ASR words vs ASR syllables
  - Significantly higher MAP on ASR words
- Multiple ASR hypotheses (words)
  - Not showing improvement of MAP
- Julius vs KALDI
  - Significantly higher MAP on KALDI 1-best words

Query-by-Query Comparison



Runs: 1-best Julius ASR words vs manual transcription

- Among 29 queries that have an AP of >=0.2 on manual transcription (reference run), 16 queries achieved only less than 20% AP of the reference run on 1-best Julius ASR words
- Detailed analysis is needed for these 16 queries/topics

Conclusions and Future Work

- Noisy ASR text degrades the retrieval effectiveness of spontaneous spoken content
  - But systems can produce ASR texts leading to results comparable to those of manual transcriptions
  - Comparative analysis of ASR texts by different systems are needed
- ASR syllables alone are not reliable for SCR
  - Maybe should better be combined with words
- Treating multiple ASR hypotheses as independent terms does not help improve retrieval effectiveness
  - More sophisticated term weighting techniques might work
- Failure analysis on problematic queries shall tell more about both ASR and IR