Leveraging Effective Query Modeling Techniques for Speech Recognition and Summarization

Kuan-Yu Chen, Shih-Hung Liu, Berlin Chen, Ea-Ee Jan, Hsin-Min Wang, Wen-Lian Hsu, Hsin-Hsi Chen

{kychen, journey, whm, hsu}@iis.sinica.edu.tw, berlin@ntnu.edu.tw, ejan@us.ibm.com, hhchen@csie.ntu.edu.tw

Summary

- Statistical language modeling (LM) has long been an interesting yet challenging research area
- LM for information retrieval (IR) has enjoyed remarkable empirical success
  - An emerging stream is to employ the pseudo-relevance feedback process to enhance the representation of the input query
- This paper presents a continuation of such a general line of research and the main contribution is three-fold
  1) We propose a principled framework which can unify the relationships among several query formulations
  2) We propose an extended query modeling formulation by incorporating critical query-specific information cues to guide the model estimation
  3) We further adopt and formalize such a framework to the speech recognition and summarization tasks

Query Modeling for Information Retrieval

Relevance Modeling (RM)

- Under the notion of relevance modeling, each query Q is assumed to be associated with an unknown relevance class R_Q, and documents that are relevant to the semantic content expressed in query are samples drawn from the relevance class R_Q
- In reality, since there is no prior knowledge about R_Q, we may use the top-ranked documents $\mathbf{D}_{Top}$ to approximate the relevance class R_Q

$$P_{\mathrm{RM}}(w|Q)=\frac{\sum_{D_r\in\mathbf{D}_{Top}}P(D_r)P(w|D_r)\prod_{w'\in Q}P(w'|D_r)}{\sum_{D_r''\in\mathbf{D}_{Top}}P(D_r'')\prod_{w'\in Q}P(w'|D_r'')}$$

Simple Mixture Modeling (SMM)

- Simple mixture model (SMM) assumes that words in $\mathbf{D}_{Top}$ are drawn from a two-component mixture model:
  1) One component is the query-specific topic model $P_{\mathrm{SMM}}(w|Q)$
  2) The other is a generic background model $P(w|BG)$

$$L=\prod_{D_r\in\mathbf{D}_{Top}}\prod_{w\in V}(\alpha\cdot P_{\mathrm{SMM}}(w|Q)+(1-\alpha)\cdot P(w|BG))^{c(w,D_r)}$$

Regularized Simple Mixture Modeling (RSMM)

- Although the SMM modeling aims to extract extra word usage cues for enhanced query modeling, it may confront two intrinsic problems
  1) One is the extraction of word usage cues from $\mathbf{D}_{Top}$ is not guided by the original query
     - ◻ This would lead to a concern for SMM to be distracted from being able to appropriately model the query of interest
  2) The other is that the mixing coefficient is fixed across all top-ranked documents
     - ◻ Different documents would potentially contribute different amounts of word usage cues to the enhanced query model

$$L=\prod_{w\in V}P_{\mathrm{RSMM}}(w|Q)^{\mu\cdot P(w|Q)}\prod_{D_r\in\mathbf{D}_{Top}}\prod_{w\in V}(\alpha_{D_r}\cdot P_{\mathrm{RSMM}}(w|Q)+(1-\alpha_{D_r})\cdot P(w|BG))^{c(w,D_r)}$$

The Proposed Modeling Framework

Fundamentals

- It is obvious that the major difference among the representative query models mentioned above is how to capitalize on the set of top-ranked documents and the original query
- A principled framework can be obtained to unify all of these query models by using a generalized objective likelihood function

$$L=\prod_{w\in V}\prod_{E_i\in\mathbf{E}}\left(\sum_{M_r\in\mathbf{M}}P(w|M_r)P(M_r)\right)^{c(w,E_i)}$$
$$s.t.\ \sum_{M_r\in\mathbf{M}}P(M_r)=1$$

where **E** represents a set of observations which we want to maximize their likelihood, and **M** denotes a set of mixture components

- Based on the proposed framework, we highlight how to infer several query modeling formulations from the unified modeling:
  1) **Relevance modeling**:
     - ◻ **E** only consists of the user query
     - ◻ **M** comprises a set of document models corresponding to the top-ranked (pseudo-relevant) documents
     - ◻ Assume the document models are known
  2) **Simple mixture modeling**:
     - ◻ **M** consists of two components: one component is a generic background model and the other is an unknown query-specific topic model
     - ◻ The weight of each component is presumably fixed in advance
     - ◻ The observations are those top-ranked documents (i.e., $\mathbf{E}=\mathbf{D}_{Top}$)
  3) **Regularized simple mixture modeling**:
     - ◻ The weight of each component is required to be estimated
     - ◻ A Dirichlet prior is placed on the enhanced query model

Query-specific Mixture Modeling (QMM)

- The SMM model and the RSMM model are intended to extract useful word usage cues from $\mathbf{D}_{top}$
  - ◻ Relevant to the original query Q and external to those already captured by the generic background model
- We argue that
  1) The "generic information" should be carefully crafted for each query due to that users' information needs may be very diverse
     - ◻ To crystallize the idea, a query-specific background model $P_Q(w|\mathrm{BG})$ for each query Q can be derived from $\mathbf{D}_{Top}$ directly
  2) Since the original query model $P(w|Q)$ cannot be accurately estimated, thus it may not necessarily be the best choice for use in defining a conjugate Dirichlet prior
     - ◻ We propose to use the RM model as a prior to guide the estimation of the enhanced query model

$$L=\prod_{w\in V}P_{\mathrm{QMM}}(w|Q)^{\mu\cdot P_{\mathrm{RM}}(w|Q)}\times$$
$$\prod_{D_r\in\mathbf{D}_{Top}}\prod_{w\in V}(\alpha_{D_r}\cdot P_{\mathrm{QMM}}(w|Q)+(1-\alpha_{D_r})\cdot P_Q(w|BG))^{c(w,D_r)}$$

Experiments

Query Modeling for Speech Recognition

- Language modeling is a critical and integral component in any large vocabulary continuous speech recognition (LVCSR) system
- The role of language modeling in LVCSR can be interpreted as calculating the conditional probability $P(w|H)$, in which H is a search history, usually expressed as a sequence of words $H=h_1, h_2,\dots, h_L$, and w is one of its possible immediately succeeding words
- For a search history H, we can conceptually regard it as a query and each of its immediately succeeding words w as a (single-word) document
- We notice three particularities from the experimental results
  - ◻ There is more fluctuation in the CER results of SMM than RM
  - ◻ The other interesting observation is that RSMM only achieves a comparable (even worse) result when compared to SMM
  - ◻ It is evident that the proposed QMM is the best-performing method among all the query models compared in the paper

	16	32	64	128
Baseline	20.08			
Cache	19.86			
LDA	19.29	19.30	19.28	19.15
RM	19.26	19.26	19.26	19.26
SMM	19.19	19.00	19.14	19.10
RSMM	19.18	19.14	19.15	19.19
QMM	19.05	18.97	19.00	18.99

Query Modeling for Speech Summarization

- Extractive speech summarization aims at producing a concise summary by selecting salient sentences or paragraphs from the original spoken document
- This task could be framed as an ad-hoc IR problem
  - ◻ The spoken document is treated as an information need
  - ◻ Each sentence of the document is regarded as a candidate information unit to be retrieved
- Two noteworthy observations can be drawn from the results
  - ◻ All these query models can considerably improve the summarization performance of the KLM (baseline) method
  - ◻ QMM is the best-performing one among all the formulations studied in this paper for both the TD and SD cases

	Manual Transcripts (TD)			ASR Transcripts (SD)		
	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L
VSM	0.347	0.228	0.290	0.342	0.189	0.287
MMR	0.407	0.294	0.358	0.381	0.226	0.331
KLM	0.411	0.298	0.361	0.364	0.210	0.307
RM	0.453	0.335	0.403	0.382	0.239	0.331
SMM	0.439	0.320	0.388	0.383	0.229	0.327
RSMM	0.472	0.365	0.423	0.381	0.235	0.329
QMM	0.486	0.382	0.435	0.395	0.256	0.349