Learning to Model Domain-Specific Utterance Sequences for Extractive Summarization of Contact Center Dialogues

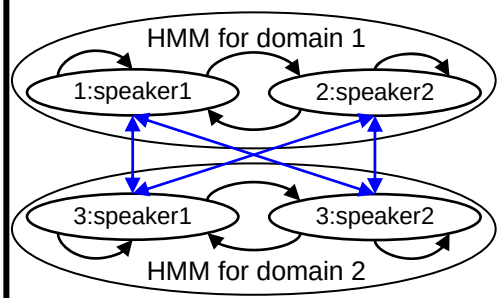
Ryuichiro Higashinaka, Yasuhiro Minami, Hitoshi Nishikawa,
Kohji Dohsaka, Toyomi Meguro, Satoshi Takahashi, Genichiro Kikui
NTT Corporation, Japan

Overview

- We propose **a novel extractive summarization method for multi-domain contact center dialogues.**
- We use **Class-Speaker Hidden Markov Models (CSHMMs)** to simultaneously model domain-specific utterance sequences and common (domain-wide) utterance sequences from multi-domain data.
  - A CSHMM is basically a concatenation of HMMs trained for each dialogue domain.
- For a dialogue in Domain k, extractive summarization is done by selecting utterances that are most likely to have been generated from the HMM for Domain k using Viterbi decoding.
- We applied CSHMMs to **contact center dialogue transcripts** of six different domains.
- **Our method outperformed competitive baselines** based on the maximum coverage of important words.

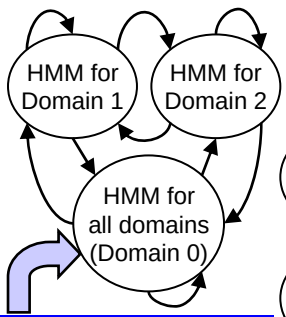
Three Topologic Variations of CSHMM

(1) Ergodic



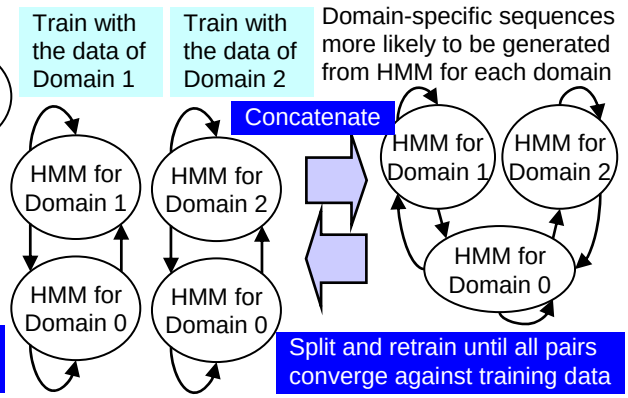
All states are connected ergodically with equal transition probabilities

(2) Ergodic with common states



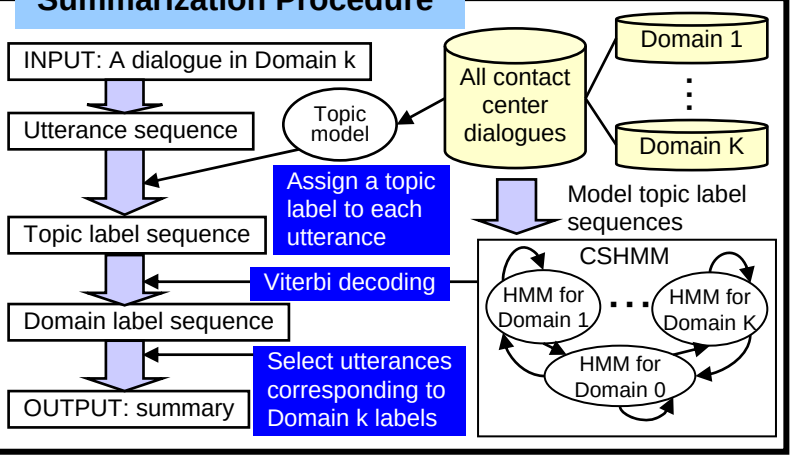
HMM (common states) trained from all domains

(3) Concatenated training



Split and retrain until all pairs converge against training data

Summarization Procedure



Dialogue Data

- **Simulated contact center dialogues in six different domains (698 dialogues)**
  - (1) Finance
  - (2) Internet service provider
  - (3) Local government unit
  - (4) Mail order
  - (5) PC support
  - (6) Telecommunication
- **Callers made calls to operators based on scenarios (description of what to ask)**
- **Training data: 391 dialogues**
- **Test data: 307 dialogues**

Evaluation Results

Our method cannot alter compression rates
➔We made our method output summaries first and made baselines output summaries with the same lengths.

- Baseline1: Maximum coverage of important words in a dialogue (importance is estimated by TF) by ILP
- Baseline2: Same as Baseline1 but the importance of words is estimated by how much each word is related to the target domain

Evaluation measure: **F-measure** (accuracy of correctly retrieving content words in the scenarios)

	Ergodic	+ Common states	Concatenated training
Proposed	0.177	0.177	0.199
Baseline1	0.171	0.171	0.163
Baseline2	0.189	0.189	0.187
Comp. rate	0.42	0.42	0.30

- **CSHMM with concatenated training significantly outperformed others ($p<0.01$).**
- **CSHMM successfully discriminates domain-specific sequences in multi-domain data.**