Score Distribution Based Term Specific Thresholding for Spoken Term Detection

D. Can  M. Saraçlar

Boğaziçi University
Department of Electrical & Electronics Engineering
BUSIM Lab
Introduction

Thresholding for Spoken Term Detection

Global Thresholding
Term Weighted Value Based Term Specific Thresholding
Score Distribution Based Term Specific Thresholding

Experiments

Setup
Results
Application: Sign Dictionary
Anatomy of a Spoken Term Detection (STD) System

User

Query

Preprocess

Retrieval

Search Engine

Speech Database

ASR

Index

Can, Saraçlar

Score Distribution Based Term Specific Thresholding for STD
Anatomy of a Spoken Term Detection (STD) System

- **User**
  - **Query**
    - **Preprocess**
      - **Search Engine**
        - **Retrieve**
          - yes
            - **Speech Database**
              - **ASR**
                - **Index**
                - **INDEXING**
              - no
                - **Dispose**

- **Can, Saracaş**
  - Score Distribution Based Term Specific Thresholding for STD
Anatomy of a Spoken Term Detection (STD) System

- **User**
- **Query**
- **Preprocess**
- **Search Engine**
- **Index**
- **Speech Database**

Decision Diagram:
- User inputs a query.
- Preprocess step.
- Search Engine queries Speech Database.
- Searching depends on whether the term is larger than $\tau$.
- If yes, search continues; if no, dispose.

Score Distribution Based Term Specific Thresholding for STD

Can, Saraçlar
Anatomy of a Spoken Term Detection (STD) System

- User
- Query
- Preprocess
- Search Engine
- Index
- Speech Database

RETRIEVAL

larger than $\tau$?

yes

no

Dispose

Can, Saraçlar
Score Distribution Based Term Specific Thresholding for STD
Outline

1. Introduction

2. Thresholding for Spoken Term Detection
   - Global Thresholding
   - Term Weighted Value Based Term Specific Thresholding
   - Score Distribution Based Term Specific Thresholding

3. Experiments
   - Setup
   - Results
Global Thresholding

- Pick a global threshold $\theta$ for all query terms
- Apply binary thresholding
- Vary $\theta$ for different operating points

No term specific behavior, no joint processing of candidates, hence poor performance!
**Global Thresholding**

- **Global Thresholding**
  - Pick a global threshold $\theta$ for all query terms
  - Apply binary thresholding
  - Vary $\theta$ for different operating points

No term specific behavior, no joint processing of candidates, hence poor performance!
**Global Thresholding**

- **Global Thresholding**
  - Pick a global threshold $\theta$ for all query terms
  - Apply binary thresholding
  - Vary $\theta$ for different operating points

**Normalized histogram of posterior scores for an example query**

No term specific behavior, no joint processing of candidates, hence poor performance!
Global Thresholding

- Pick a global threshold $\theta$ for all query terms
- Apply binary thresholding
- Vary $\theta$ for different operating points

No term specific behavior, no joint processing of candidates, hence poor performance!
Global Thresholding

- Pick a global threshold $\theta$ for all query terms
- Apply binary thresholding
- Vary $\theta$ for different operating points

No term specific behavior, no joint processing of candidates, hence poor performance!
Outline

1 Introduction

2 Thresholding for Spoken Term Detection
   - Global Thresholding
   - Term Weighted Value Based Term Specific Thresholding
   - Score Distribution Based Term Specific Thresholding

3 Experiments
   - Setup
   - Results
Term Weighted Value (TWV) [NIST, 2006]

\[
TWV = 1 - \frac{1}{Q} \sum_{k=1}^{Q} \{ P_{\text{miss}}(q_k) + \beta P_{\text{FA}}(q_k) \}
\]

\[
P_{\text{miss}}(q_k) = 1 - \frac{C(q_k)}{R(q_k)}, \quad P_{\text{FA}}(q_k) = \frac{A(q_k) - C(q_k)}{T - C(q_k)}
\]

- \( Q \) Number of queries
- \( R(q_k) \) Number of occurrences of query \( q_k \)
- \( A(q_k) \) Total number of retrieved documents for \( q_k \)
- \( C(q_k) \) Number of correctly retrieved documents for \( q_k \)
- \( T \) Total duration of the speech archive
- \( \beta \) Cost of false alarms relative to hits
TWV Based Term Specific Thresholding [Miller et al., 2007]

\[
\hat{V}_{hit}(q_k) = \frac{1}{\hat{N}_{true}(q_k)}, \quad \hat{C}_{FA}(q_k) = \frac{\beta}{T - \hat{N}_{true}(q_k)}
\]

\[
\hat{\theta}(q_k) = \frac{\hat{C}_{FA}(q_k)}{\hat{C}_{FA}(q_k) + \hat{V}_{hit}(q_k)}
\]

\[\hat{N}_{true}(q_k)\] Expected count of occurrences of \(q_k\)
\[\hat{\theta}(q_k)\] Optimal threshold for \(q_k\) maximizing TWV in the expected sense

- Term specific expected counts → Term specific thresholds
- Vary \(\beta\) for different operating points

Only the sum of individual scores affects the threshold!
TWV Based Term Specific Thresholding [Miller et al., 2007]

\[ \hat{V}_{hit}(q_k) = \frac{1}{\hat{N}_{true}(q_k)}, \quad \hat{C}_{FA}(q_k) = \frac{\beta}{T - \hat{N}_{true}(q_k)} \]

\[ \hat{\theta}(q_k) = \frac{\hat{C}_{FA}(q_k)}{\hat{C}_{FA}(q_k) + \hat{V}_{hit}(q_k)} \]

- \( \hat{N}_{true}(q_k) \): Expected count of occurrences of \( q_k \)
- \( \hat{\theta}(q_k) \): Optimal threshold for \( q_k \) maximizing TWV in the expected sense

- Term specific expected counts \( \rightarrow \) Term specific thresholds
- Vary \( \beta \) for different operating points

Only the sum of individual scores affects the threshold!
Outline

1. Introduction

2. Thresholding for Spoken Term Detection
   - Global Thresholding
   - Term Weighted Value Based Term Specific Thresholding
   - Score Distribution Based Term Specific Thresholding

3. Experiments
   - Setup
   - Results

Score Distribution Based Term Specific Thresholding for STD
Scores follow exponential-like distributions

Model both classes \((c_0, c_1)\) with exponential distributions:

\[
p(x|c_0) = \lambda_0 e^{-\lambda_0 x}
\]
\[
p(x|c_1) = \lambda_1 e^{-\lambda_1 (1-x)}
\]

Model all candidates as a mixture of exponentials

\[
p(x) = P(c_0)p(x|c_0) + P(c_1)p(x|c_1)
\]

Use EM to estimate parameters \((\lambda_0, \lambda_1, P(c_0), P(c_1))\)
Exploiting Score Distributions [Manmatha et al., 2001]

- Scores follow exponential-like distributions
- Model both classes \((c_0, c_1)\) with exponential distributions:
  
  \[
  p(x|c_0) = \lambda_0 e^{-\lambda_0 x} \\
  p(x|c_1) = \lambda_1 e^{-\lambda_1 (1-x)}
  \]

- Model all candidates as a mixture of exponentials
  
  \[
  p(x) = P(c_0)p(x|c_0) + P(c_1)p(x|c_1)
  \]

- Use EM to estimate parameters \((\lambda_0, \lambda_1, P(c_0), P(c_1))\)
Computing Term Specific Thresholds

**Cost Scheme**

\[
\gamma_c(d) = \begin{cases} 
1 & d = 0, c \in c_1 \\
\alpha & d = 1, c \in c_0 
\end{cases}
\]

where \( c \) is a candidate, \( d \) is a decision, and \( \alpha \) is the cost of false alarms relative to hits.

- Estimate mixture parameters \( \rightarrow \) each component \( \sim \) a class, mixture weights \( \sim \) priors
- Bayes-optimal threshold \( \theta \) is given as:

\[
\theta = \frac{\lambda_1 + \log(\lambda_0/\lambda_1) + \log(P(c_0)/P(c_1)) + \log \alpha}{\lambda_0 + \lambda_1}.
\]

- Different operating points can be achieved by changing \( \alpha \).
Outline

1 Introduction

2 Thresholding for Spoken Term Detection
   - Global Thresholding
   - Term Weighted Value Based Term Specific Thresholding
   - Score Distribution Based Term Specific Thresholding

3 Experiments
   - Setup
   - Results
**Experimental Setup**

**Query Set**

10229 single word queries selected from Turkish Radio and Television Channel 2 (TRT2) hearing impaired news

**LVCSR System**

- **Acoustic Data:** 111 hours of BN data
  - **Train set:** 100 hours (from various TV and radio broadcasts)
  - **Test set:** 11 hours (from TRT2 hearing impaired news)

- **Language Data:** 100M words from various text sources

- **IBM Attila Speech Recognition Toolkit**
  - Baseline MLE models
  - WER on the test set: 17%
Outline

1. Introduction

2. Thresholding for Spoken Term Detection
   - Global Thresholding
   - Term Weighted Value Based Term Specific Thresholding
   - Score Distribution Based Term Specific Thresholding

3. Experiments
   - Setup
   - Results
Experimental Results

Better performance in the high precision region

Large room for improvement

Can, Saraclar
Score Distribution Based Term Specific Thresholding for STD
Experimental Results

Better performance in the high precision region

Large room for improvement

- Global Thresholding
- Term Specific Thresholding (TWV)
- EMM + EM + MBR Detection
- Cheat + EMM + MBR Detection

Score Distribution Based Term Specific Thresholding for STD
Experimental Results

Better performance in the high precision region

Large room for improvement

- Global Thresholding
- Term Specific Thresholding (TWV)
- EMM + EM + MBR Detection
- Cheat + EMM + MBR Detection
Exploiting score distributions leads to a viable term specific thresholding method.

Proposed method has a large potential as indicated by the cheat experiment.

Superior to TWV based method in the high precision region.
References I

- Manmatha, R., Rath, T., and Feng, F. (2001). Modeling score distributions for combining the outputs of search engines. In *SIGIR ’01*, pages 267–275, New York, NY, USA. ACM.

- Miller, D. R. H., Kleber, M., Kao, C., Kimball, O., Colthurst, T., Lowe, S. A., Schwartz, R. M., and Gish, H. (2007). Rapid and accurate spoken term detection. In *Proc. Interspeech*, pages 314–317.

- NIST (2006). The Spoken Term Detection (STD) 2006 Evaluation Plan http://www.nist.gov/speech/tests/std/.
EM Parameter Updates

- Model all candidates as a mixture of exponentials

\[ p(x) = P(c_0)p(x|c_0) + P(c_1)p(x|c_1) \]

- Use EM to estimate parameters \((\lambda_0, \lambda_1, P(c_0), P(c_1))\) given the scores \(x_i\) for \(i = 1, \ldots, N\).
  - First compute \(P(c_j|x_i) = P(c_j)p(x_i|c_j)/p(x_i)\) for \(j = 1, 2\)
  - Then update

\[
P(c_j) = \frac{1}{N} \sum_i P(c_j|x_i),
\]

\[
\lambda_0 = \frac{\sum_i P(c_0|x_i)}{\sum_i P(c_0|x_i)x_i},
\]

\[
\lambda_1 = \frac{\sum_i P(c_1|x_i)}{\sum_i P(c_1|x_i)(1-x_i)}.
\]