Automatic Key Term Extraction from Spoken Course Lectures

Using Branching Entropy and Prosodic/Semantic Features

Yun-Nung (**Vivian**) Chen, Yu Huang,
Sheng-Yi Kong, Lin-Shan Lee

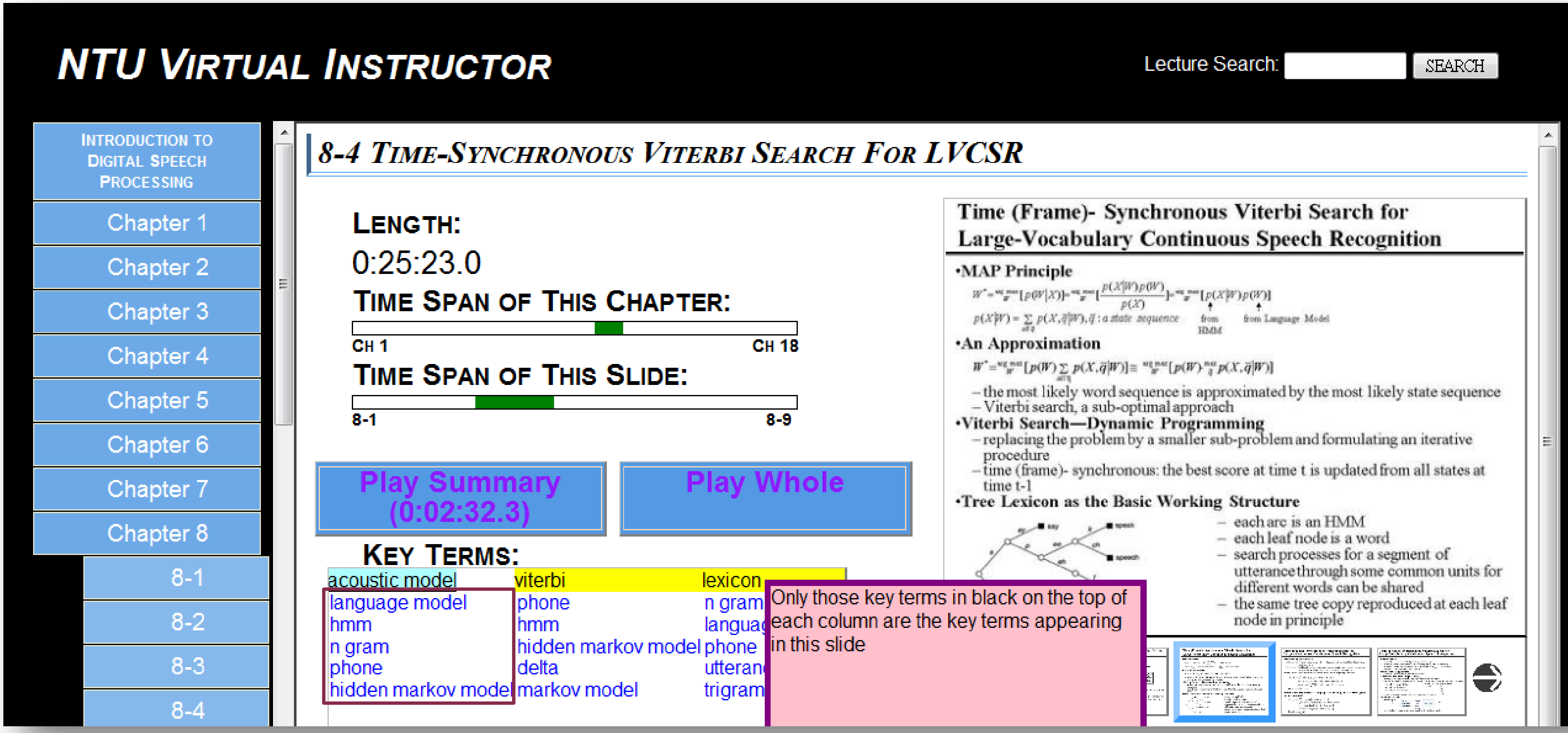
National Taiwan University, Taiwan

Introduction

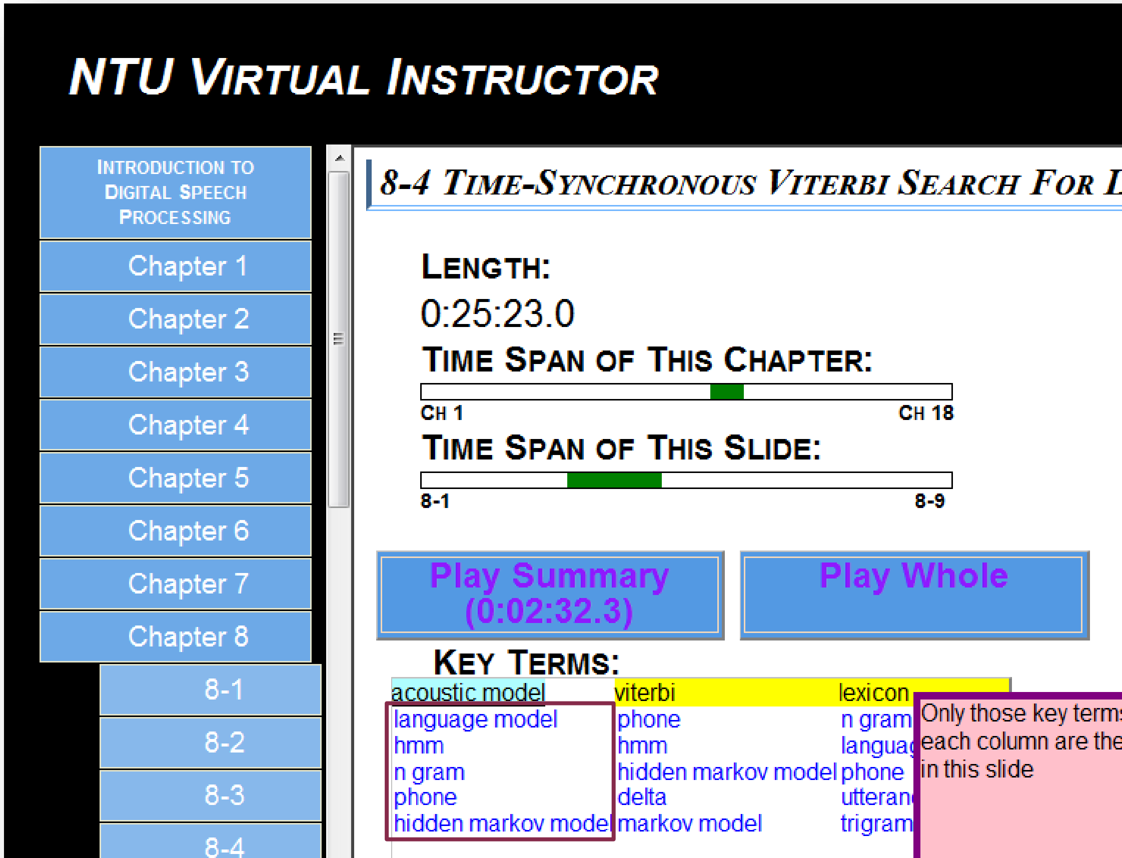
Definition

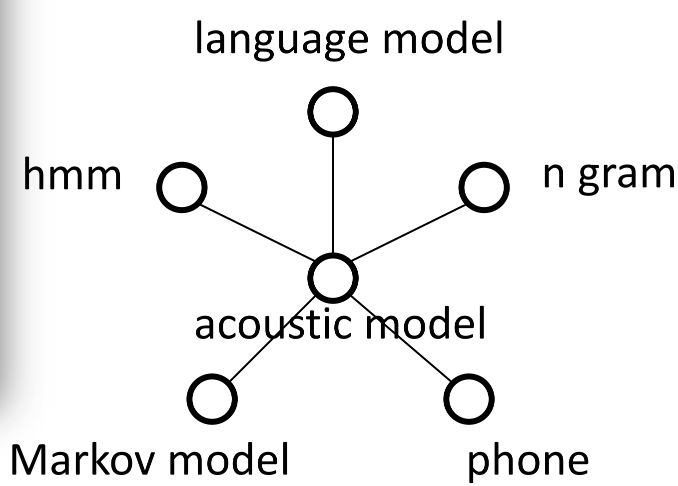
- Key Term
  - Higher term frequency
  - Core content
- Two types
  - Keyword
  - Key phrase
- Advantage
  - Indexing and retrieval
  - The relations between key terms and segments of documents

Introduction

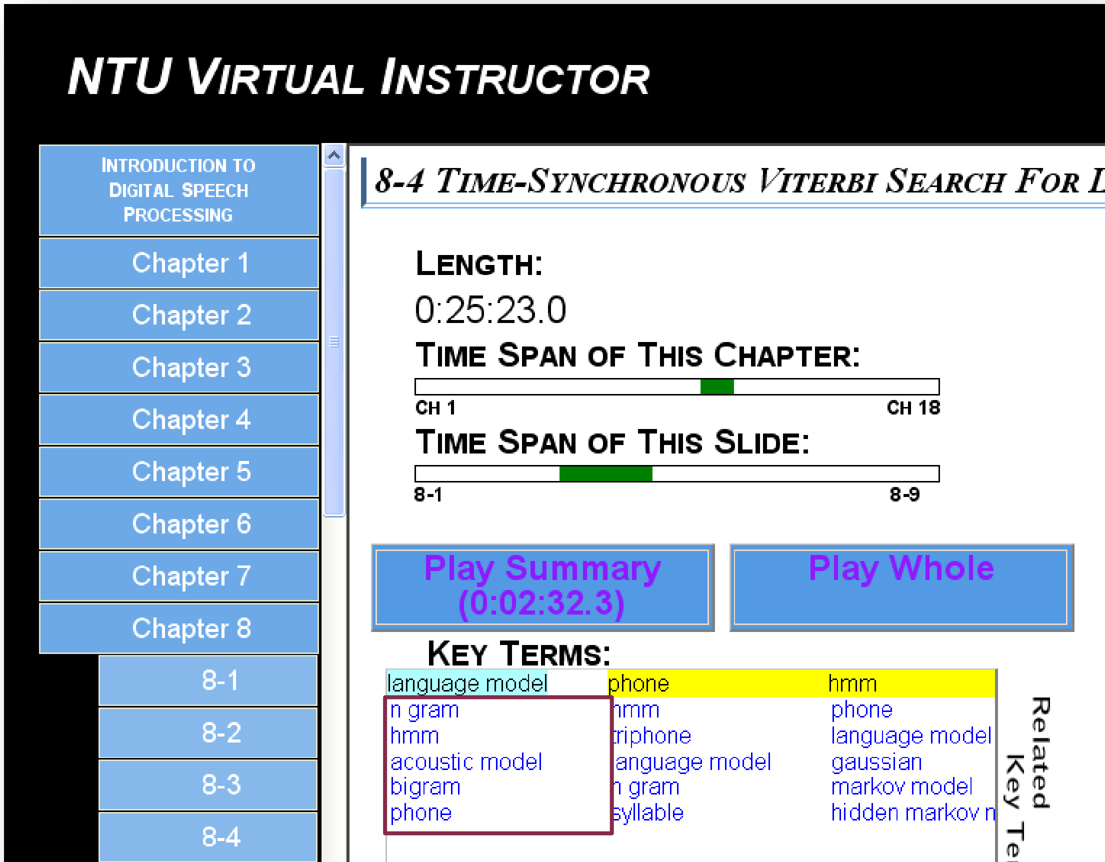


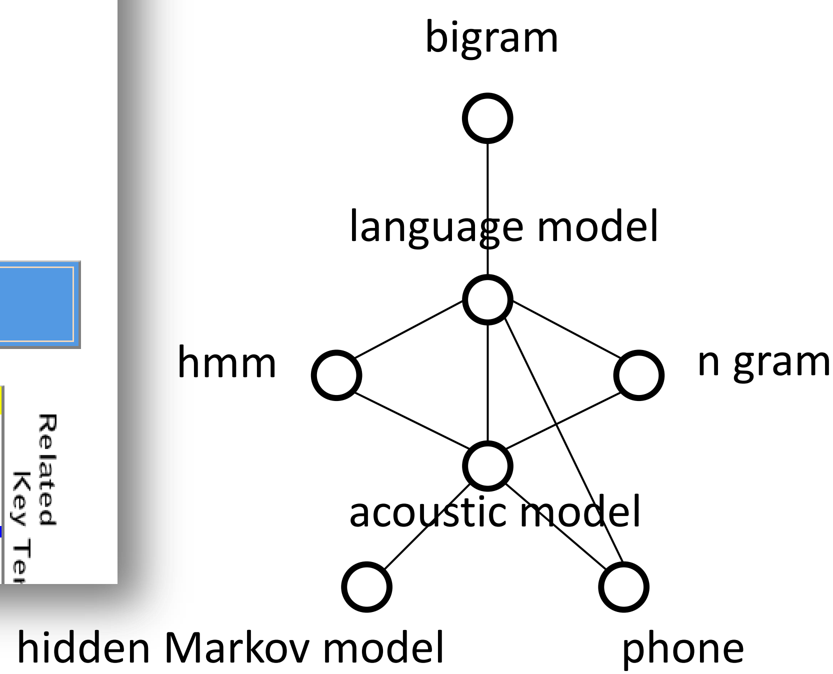
Introduction





Introduction



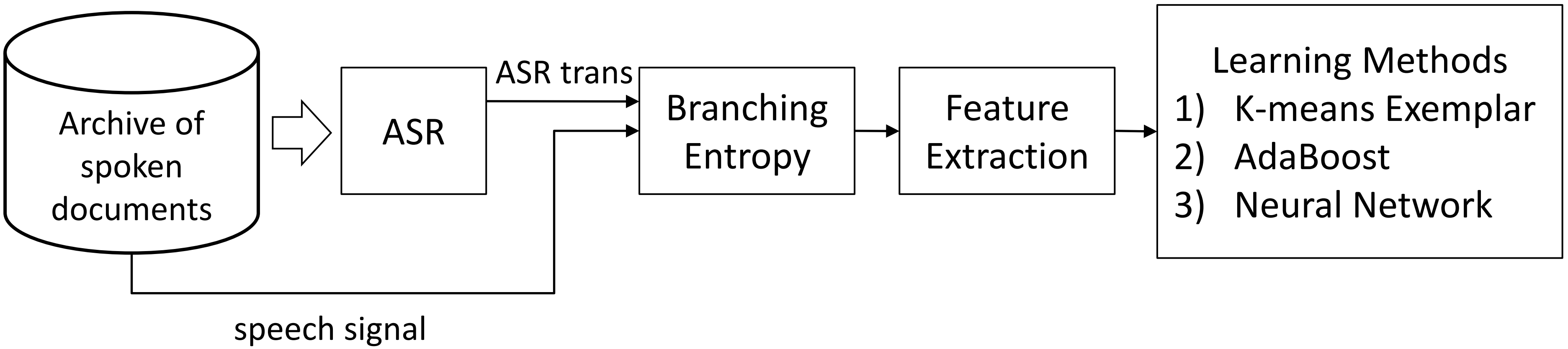


Target: extract key terms from course lectures

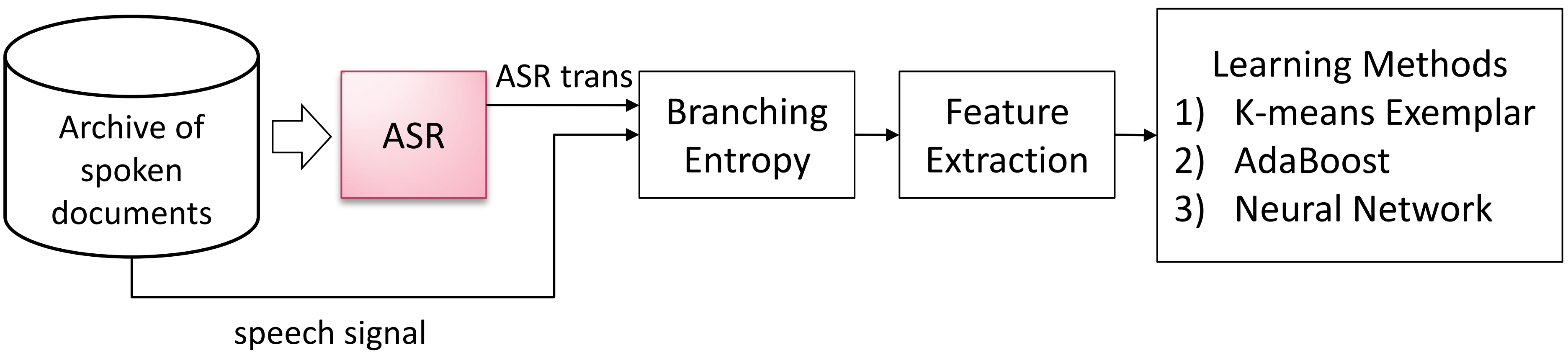
Proposed Approach

Automatic Key Term Extraction

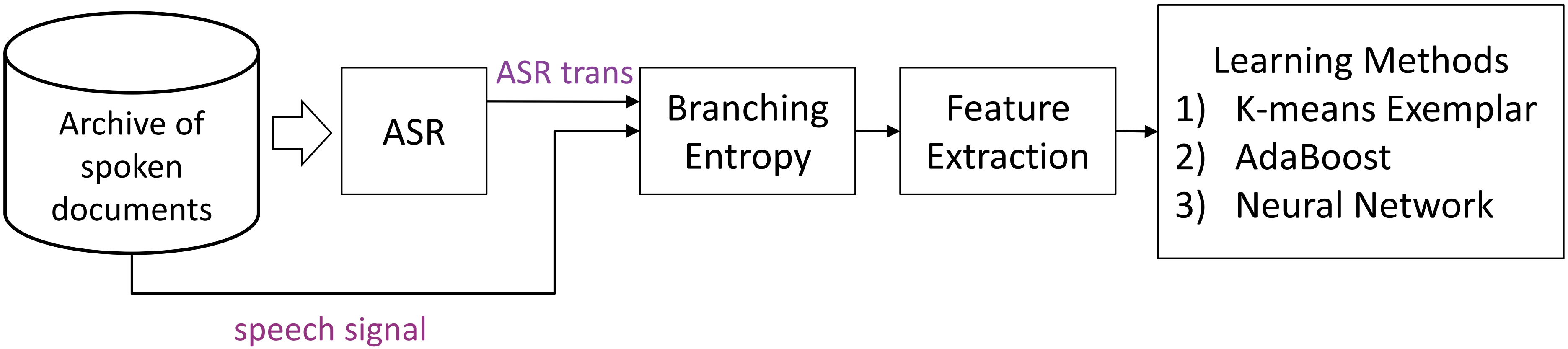
▼ Original spoken documents



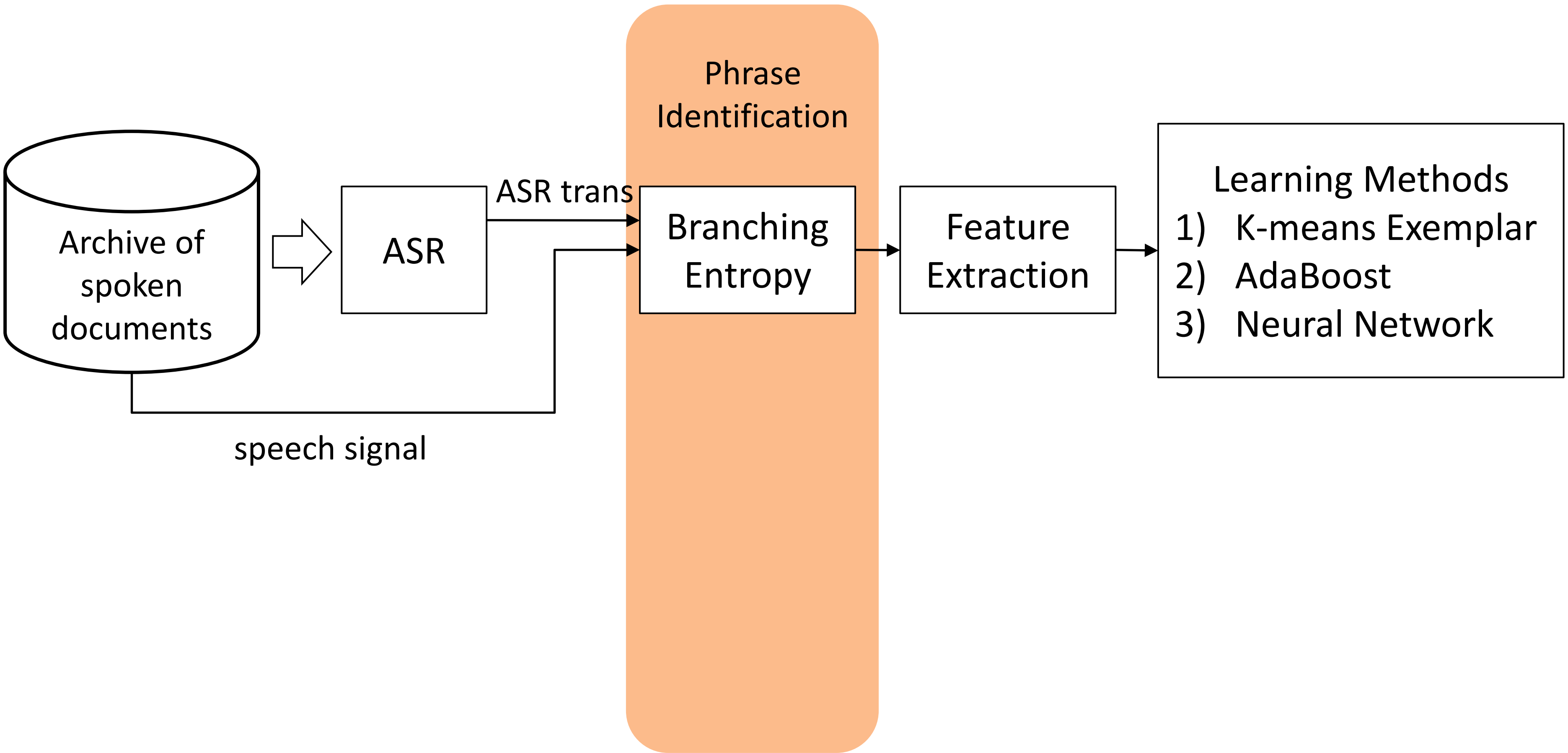
Automatic Key Term Extraction



Automatic Key Term Extraction

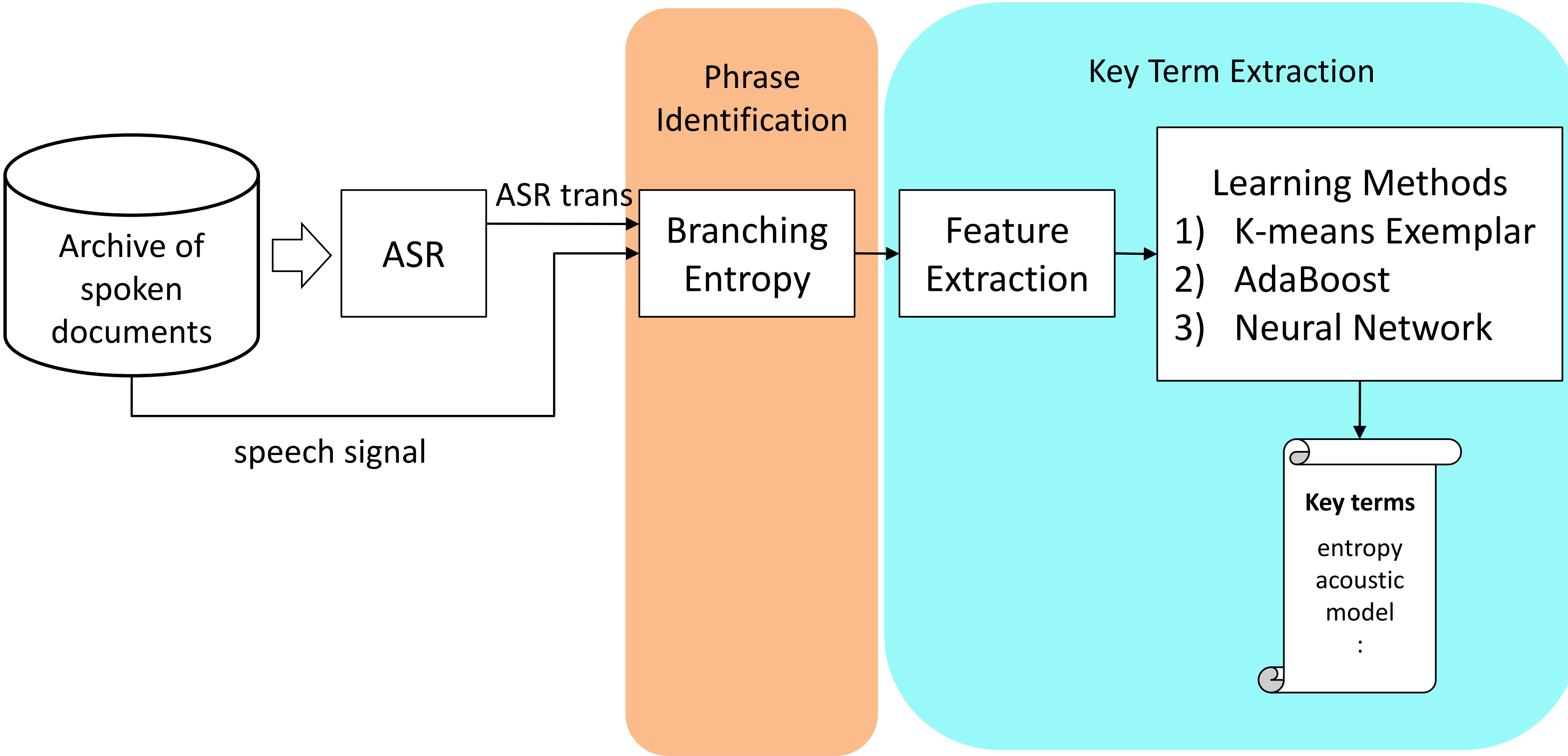


Automatic Key Term Extraction



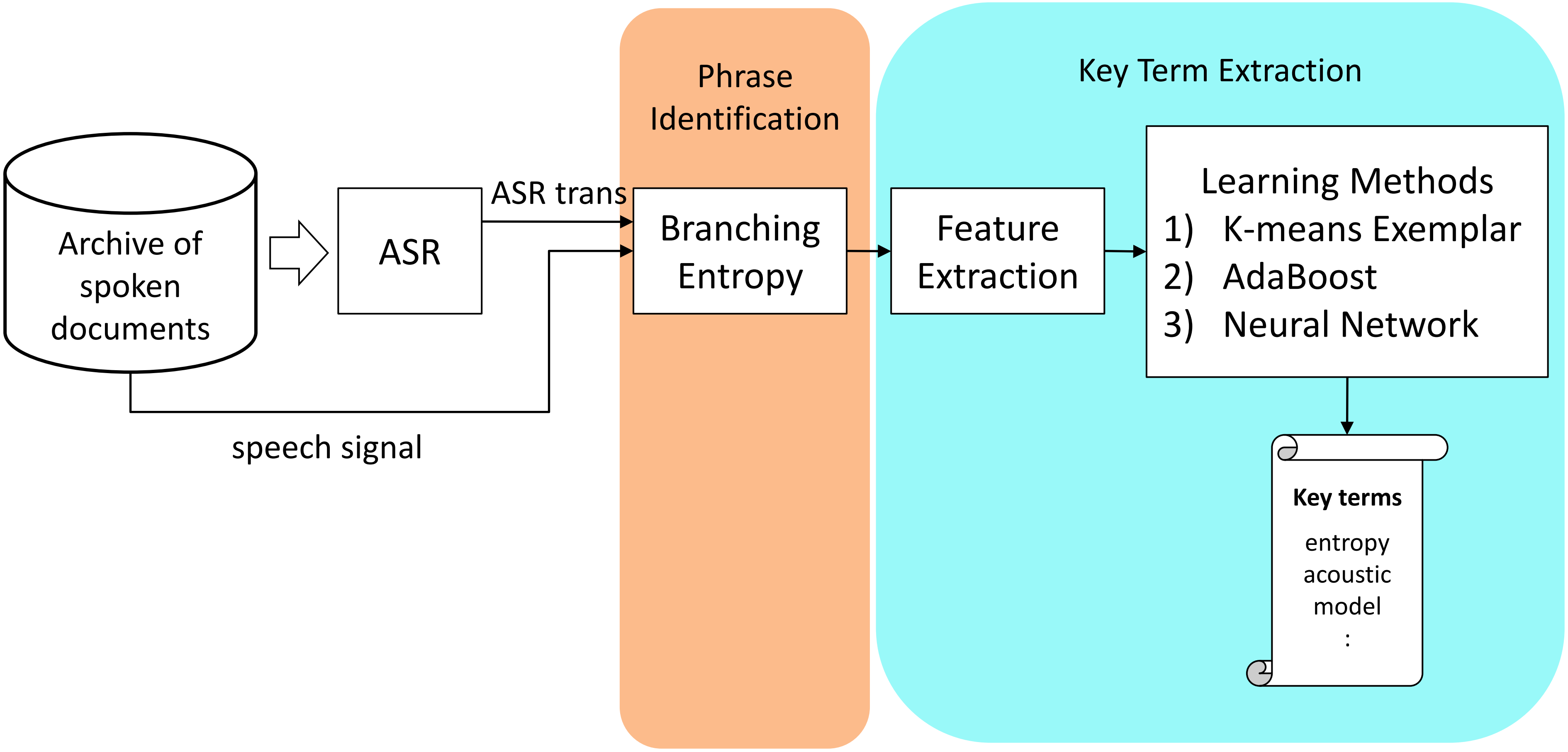
First using branching entropy to identify phrases

Automatic Key Term Extraction



Then using learning methods to extract key terms by some features

Automatic Key Term Extraction



Branching Entropy

How to decide the boundary of a phrase?

- "hidden" is almost always followed by the same word

Branching Entropy

How to decide the boundary of a phrase?

- “hidden” is almost always followed by the same word
- “hidden Markov” is almost always followed by the same word

Branching Entropy

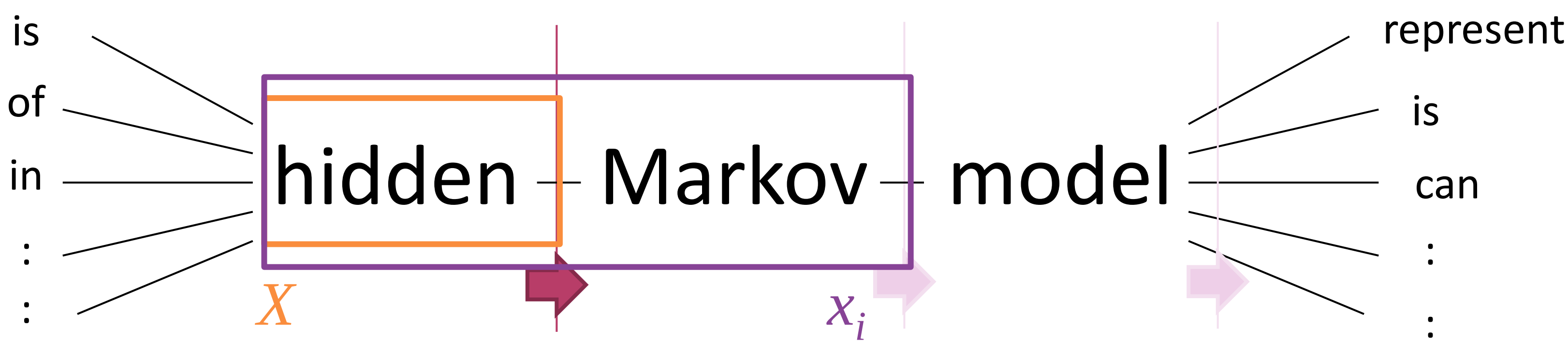
How to decide the boundary of a phrase?

- "hidden" is almost always followed by the same word
- "hidden Markov" is almost always followed by the same word
- "hidden Markov model" is followed by many different words

Define branching entropy to decide possible boundary

Branching Entropy

How to decide the boundary of a phrase?



- Definition of Right Branching Entropy
  - Probability of children x_i for X

$$p(x_i) = \frac{f_{x_i}}{f_X} \quad \begin{array}{l} X: w_1 \ldots w_k \\ x_i: w_1 \ldots w_k w^i_{(k+1)} \end{array}$$

  - Right branching entropy for X

$$H_r(X) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i)$$

Branching Entropy

How to decide the boundary of a phrase?

- Decision of Right Boundary
  - Find the right boundary located between X and x_i where

$$H_r(X) > \text{average } H_r(X)$$

Branching Entropy

How to decide the boundary of a phrase?

Branching Entropy

How to decide the boundary of a phrase?

Branching Entropy

How to decide the boundary of a phrase?

Branching Entropy

How to decide the boundary of a phrase?

- Decision of Left Boundary
  - Find the left boundary located between $\bar{X}$ and x_i where

$\bar{X}$: model Markov hidden

$$H_l(\bar{X}) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i)$$

$$H_l(\bar{X}) > \text{average } H_l(\bar{X})$$

Using PAT Tree to implement

Branching Entropy

How to decide the boundary of a phrase?

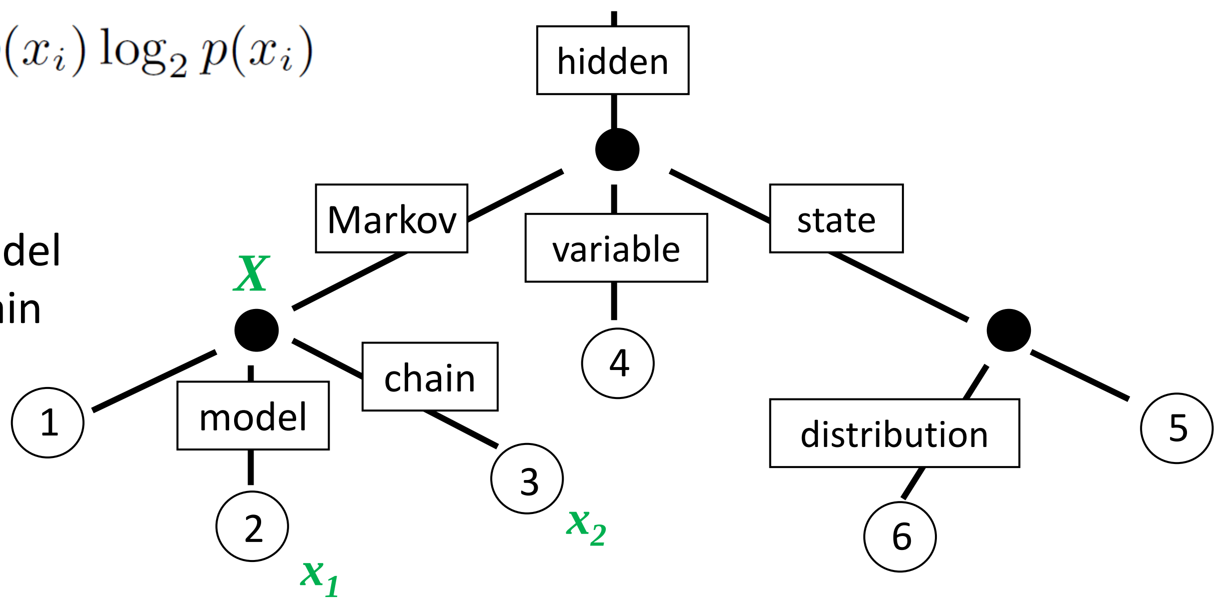
- Implementation in the PAT tree
  - Probability of children x_i for X

$$p(x_i) = \frac{f_{x_i}}{f_X} \quad \begin{array}{l} X: w_1 \ldots w_k \\ x_i: w_1 \ldots w_k w^i_{(k+1)} \end{array}$$

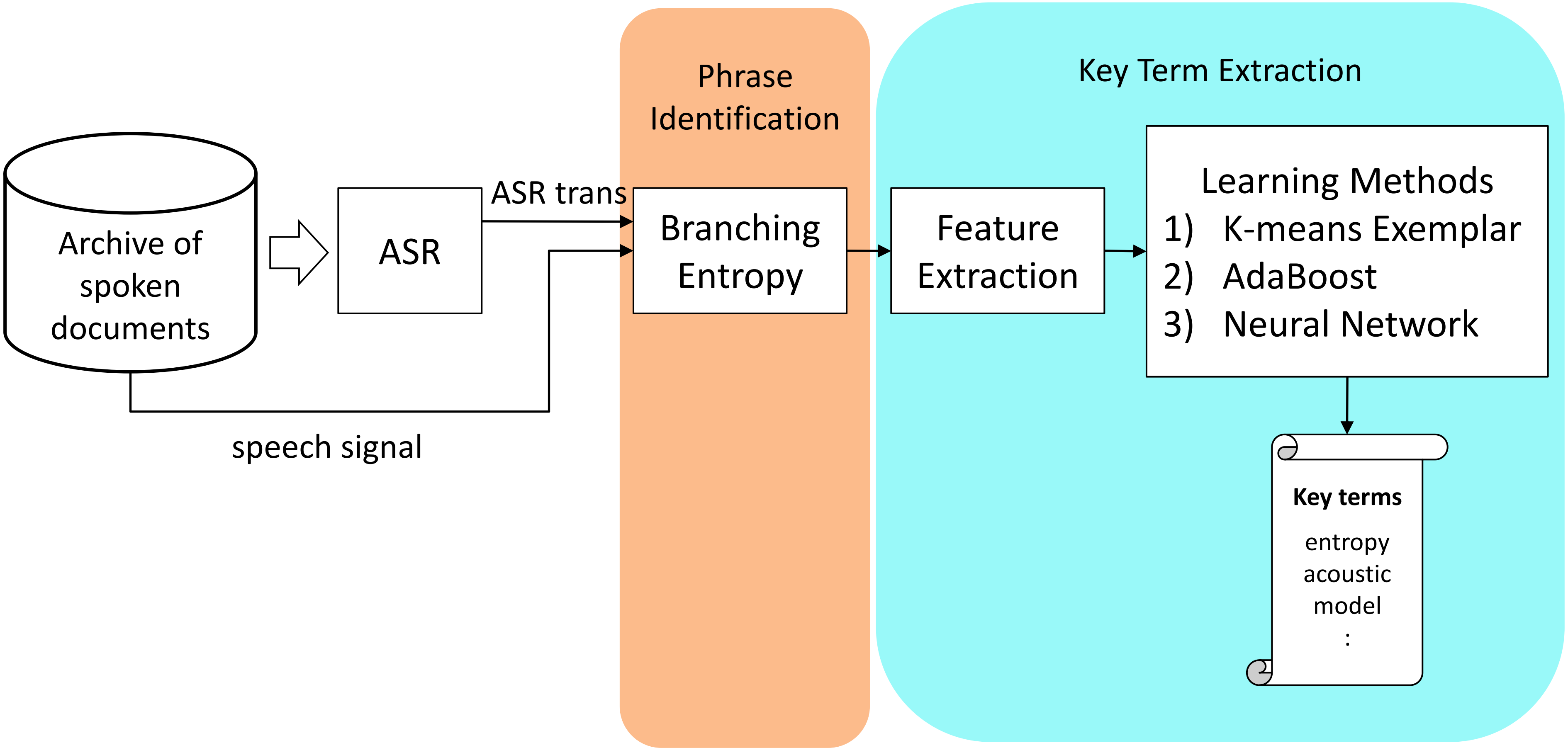
  - Right branching entropy for X

$$H_r(X) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i)$$

X : hidden Markov
x_1: hidden Markov model
x_2: hidden Markov chain



Automatic Key Term Extraction



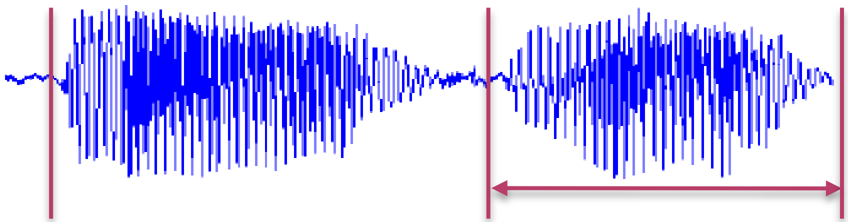
Extract prosodic, lexical, and semantic features for each candidate term

Feature Extraction

- Prosodic features

Speaker tends to use longer duration to emphasize key terms

  - For each candidate term appearing at the first time



duration of phone “a” normalized by avg duration of phone “a”

Feature Name	Feature Description
Duration (I – IV)	normalized duration (max, min, mean, range)

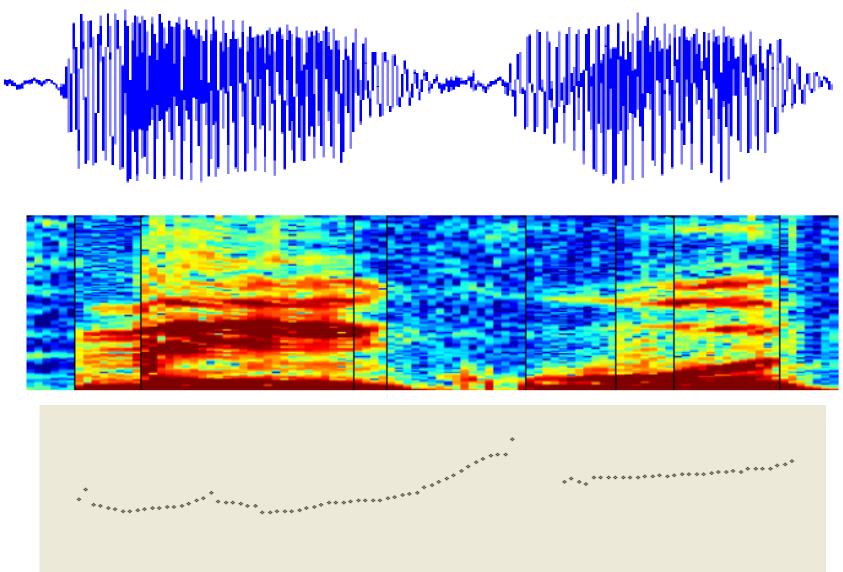
using 4 values for duration of the term

Feature Extraction

- Prosodic features

Higher pitch may represent significant information

  - For each candidate term appearing at the first time



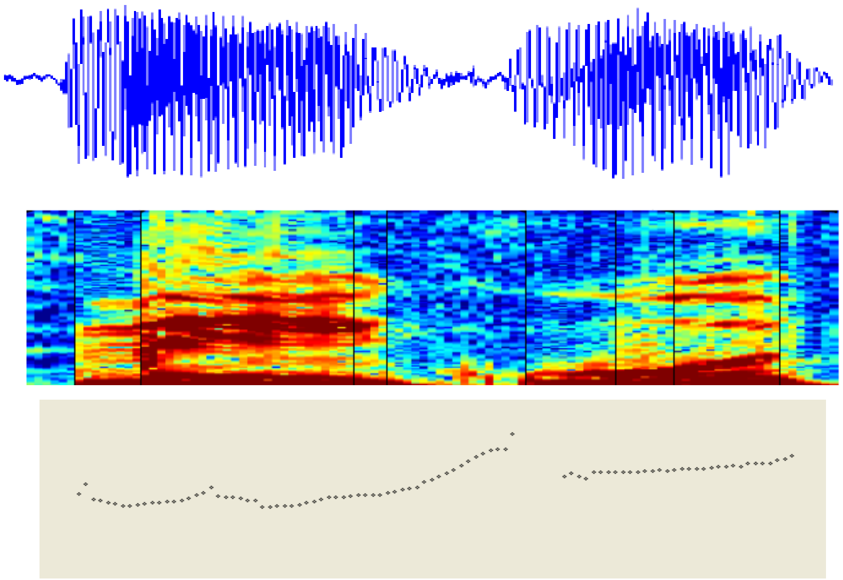
Feature Name	Feature Description
Duration (I – IV)	normalized duration (max, min, mean, range)

Feature Extraction

- Prosodic features

Higher pitch may represent significant information

- For each candidate term appearing at the first time



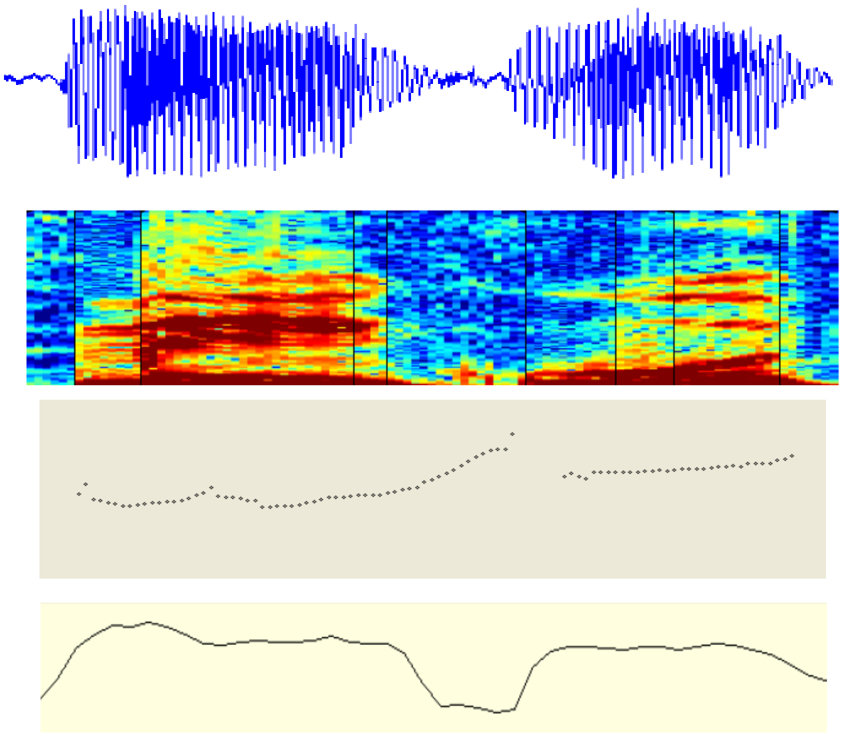
Feature Name	Feature Description
Duration (I – IV)	normalized duration (max, min, mean, range)
Pitch (I - IV)	F0 (max, min, mean, range)

Feature Extraction

- Prosodic features

Higher energy emphasizes important information

  - For each candidate term appearing at the first time



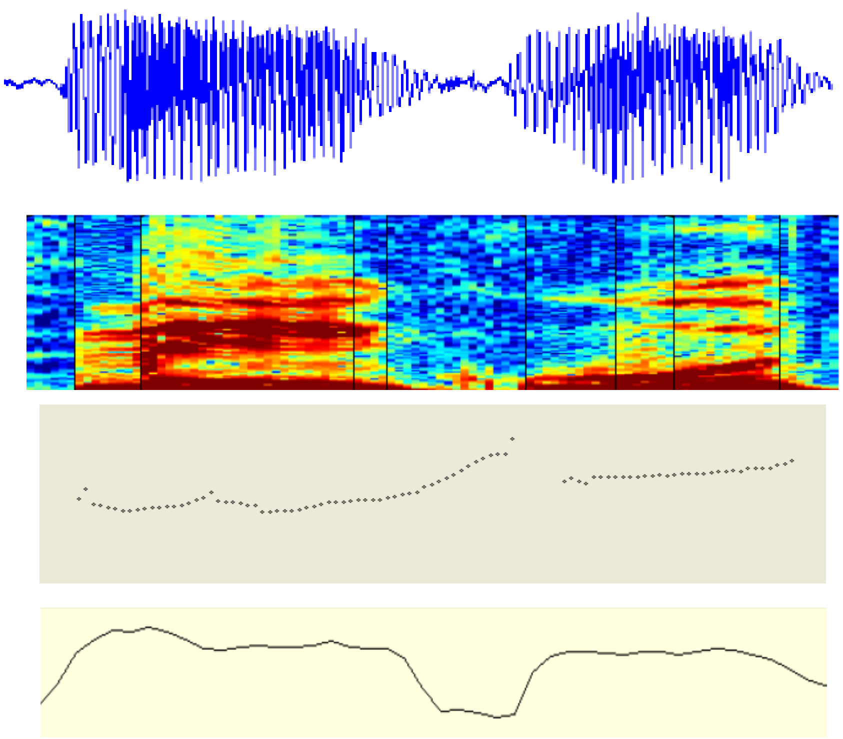
Feature Name	Feature Description
Duration (I – IV)	normalized duration (max, min, mean, range)
Pitch (I - IV)	F0 (max, min, mean, range)

Feature Extraction

- Prosodic features

Higher energy emphasizes important information

  - For each candidate term appearing at the first time



Feature Name	Feature Description
Duration (I – IV)	normalized duration (max, min, mean, range)
Pitch (I - IV)	F0 (max, min, mean, range)
Energy (I - IV)	energy (max, min, mean, range)

Feature Extraction

- Lexical features

Feature Name	Feature Description
TF	term frequency
IDF	inverse document frequency
TFIDF	tf * idf
PoS	the PoS tag

Using some well-known lexical features for each candidate term

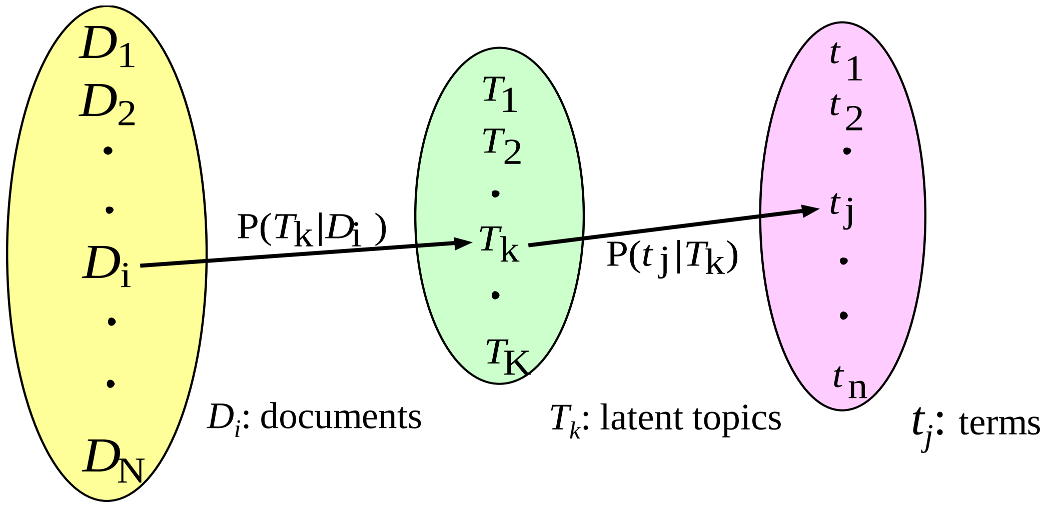
Feature Extraction

- Semantic features

Key terms tend to focus on limited topics

- Probabilistic Latent Semantic Analysis (PLSA)
  - Latent Topic Probability

$$P(T_k|t_i) = \frac{P(t_i|T_k)P(T_k)}{P(t_i)}$$



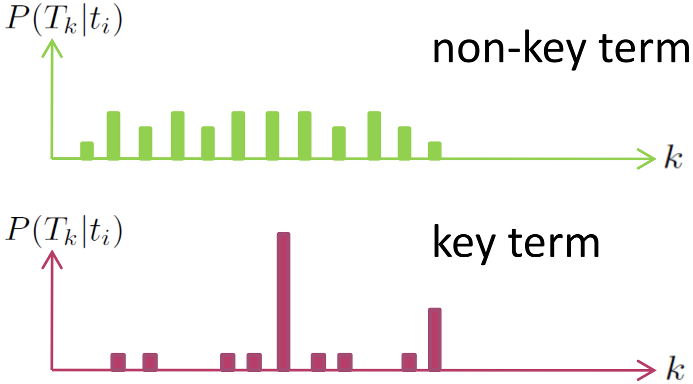
Feature Extraction

- Semantic features

Key terms tend to focus on limited topics

- Probabilistic Latent Semantic Analysis (PLSA)
  - Latent Topic Probability

$$P(T_k|t_i) = \frac{P(t_i|T_k)P(T_k)}{P(t_i)}$$



How to use it?

Feature Name	Feature Description
LTP (I - III)	Latent Topic Probability (mean, variance, standard deviation)

describe a probability distribution

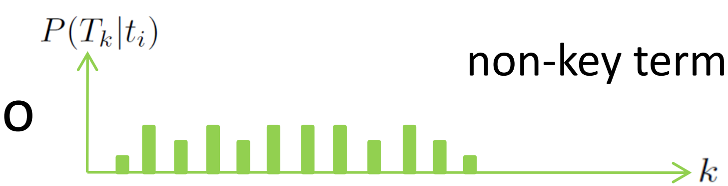
Feature Extraction

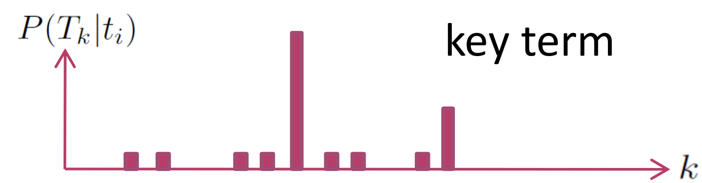
- Semantic features

Key terms tend to focus on limited topics

- Probabilistic Latent Semantic Analysis (PLSA)
  - Latent Topic Significance

    Within-topic to out-of-topic ratio





$$S_{t_i}(T_k) = \frac{\boxed{\sum_{d_j \in D} n(t_i, d_j) P(T_k|d_j)}}{\boxed{\sum_{d_j \in D} n(t_i, d_j)[1 - P(T_k|d_j)]}}$$

within-topic freq.

out-of-topic freq.

Feature Name	Feature Description
LTP (I - III)	Latent Topic Probability (mean, variance, standard deviation)

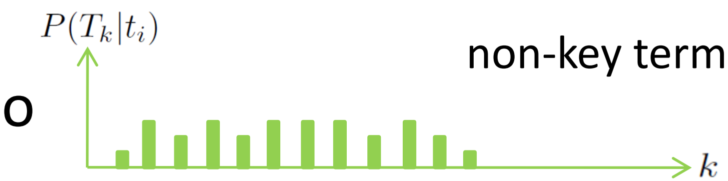
Feature Extraction

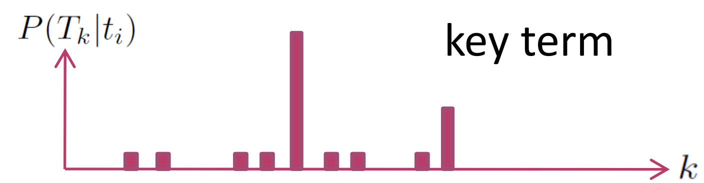
- Semantic features

Key terms tend to focus on limited topics

- Probabilistic Latent Semantic Analysis (PLSA)
  - Latent Topic Significance

    Within-topic to out-of-topic ratio





$$S_{t_i}(T_k) = \frac{\boxed{\sum_{d_j \in D} n(t_i, d_j) P(T_k|d_j)}}{\boxed{\sum_{d_j \in D} n(t_i, d_j)[1 - P(T_k|d_j)]}}$$

within-topic freq.

out-of-topic freq.

Feature Name	Feature Description
LTP (I - III)	Latent Topic Probability (mean, variance, standard deviation)
LTS (I - III)	Latent Topic Significance (mean, variance, standard deviation)

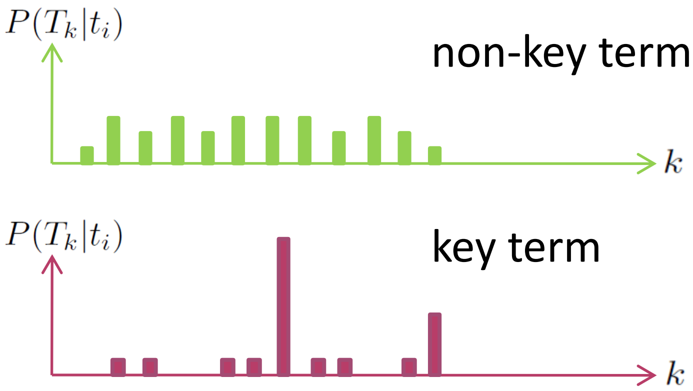
Feature Extraction

- Semantic features

Key terms tend to focus on limited topics

  - Probabilistic Latent Semantic Analysis (PLSA)
    - Latent Topic Entropy

$$EN(t_i) = -\sum_{k=1}^{K} P(T_k|t_i) \log P(T_k|t_i)$$



Feature Name	Feature Description
LTP (I - III)	Latent Topic Probability (mean, variance, standard deviation)
LTS (I - III)	Latent Topic Significance (mean, variance, standard deviation)

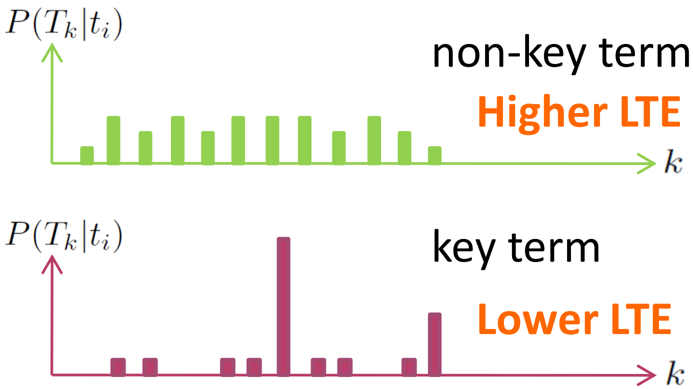
Feature Extraction

- Semantic features

Key terms tend to focus on limited topics

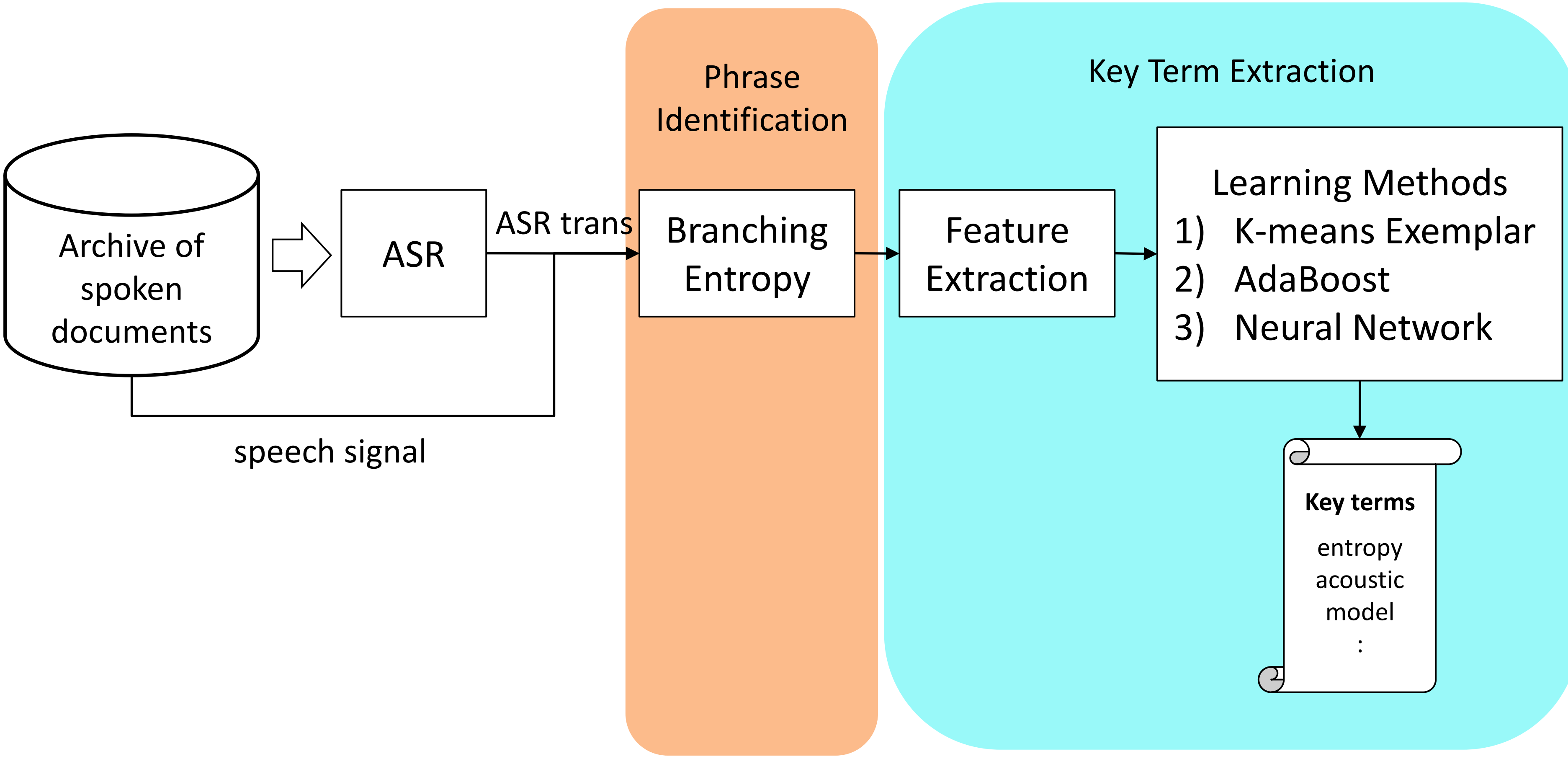
  - Probabilistic Latent Semantic Analysis (PLSA)
    - Latent Topic Entropy

$$EN(t_i) = -\sum_{k=1}^{K} P(T_k|t_i) \log P(T_k|t_i)$$



Feature Name	Feature Description
LTP (I - III)	Latent Topic Probability (mean, variance, standard deviation)
LTS (I - III)	Latent Topic Significance (mean, variance, standard deviation)
LTE	term entropy for latent topic

Automatic Key Term Extraction

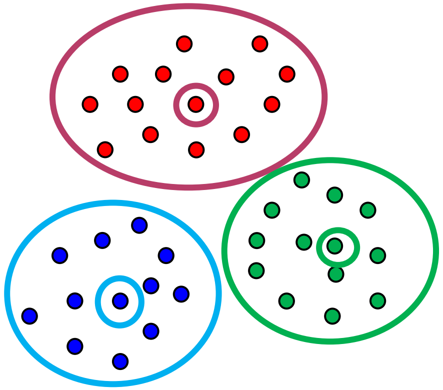


Using unsupervised and supervised approaches to extract key terms

Learning Methods

- Unsupervised learning
  - K-means Exemplar
    - Transform a term into a vector in LTS (Latent Topic Significance) space

$$v_i = (S_{t_i}(T_1), S_{t_i}(T_2), ..., S_{t_i}(T_K))$$



- Run K-means

The terms in the same cluster focus on a single topic

- Find the centroid of each cluster to be the key term

The candidate term in the same group are related to the key term
The key term can represent this topic

Learning Methods

- Supervised learning
  - Adaptive Boosting
  - Neural Network

Automatically adjust the weights of features to produce a classifier

Experiments & Evaluation

Experiments

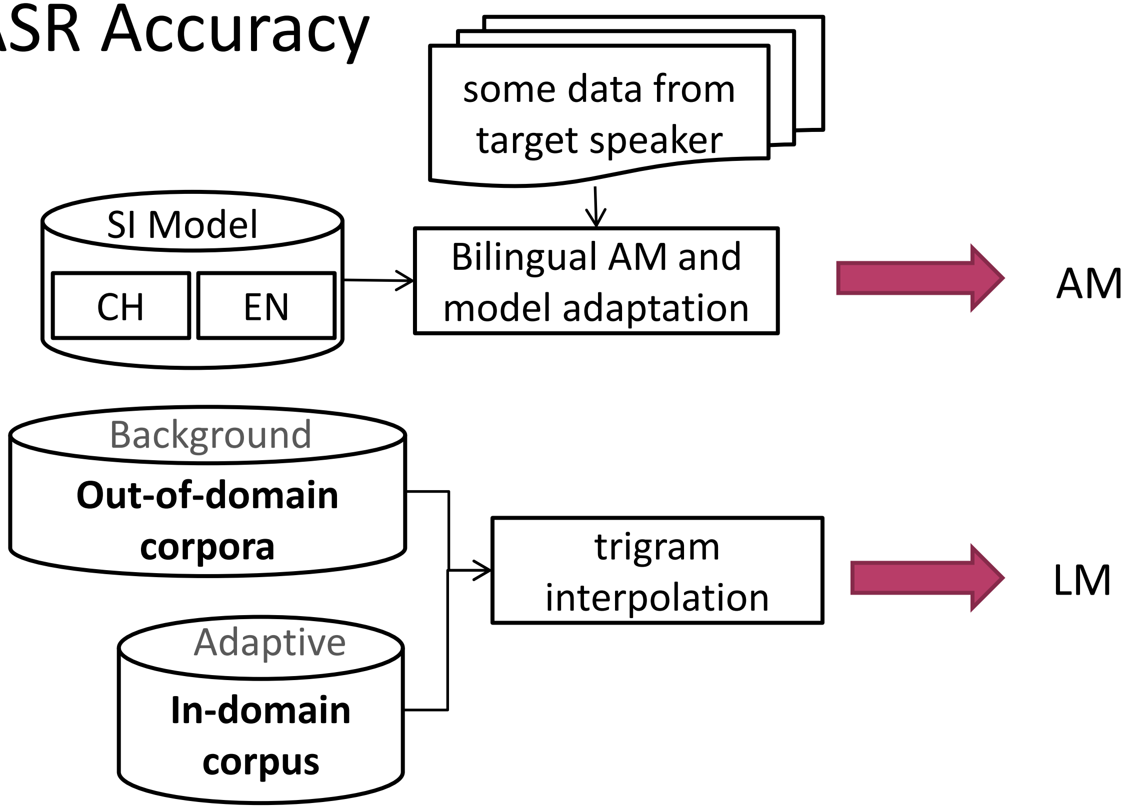
- Corpus
  - NTU lecture corpus
    - Mandarin Chinese embedded by English words

我們的solution是viterbi algorithm
(Our solution is viterbi algorithm)

    - Single speaker
    - 45.2 hours

Experiments

- ASR Accuracy



Language	Mandarin	English	Overall
Char Acc (%)	78.15	53.44	76.26

Experiments

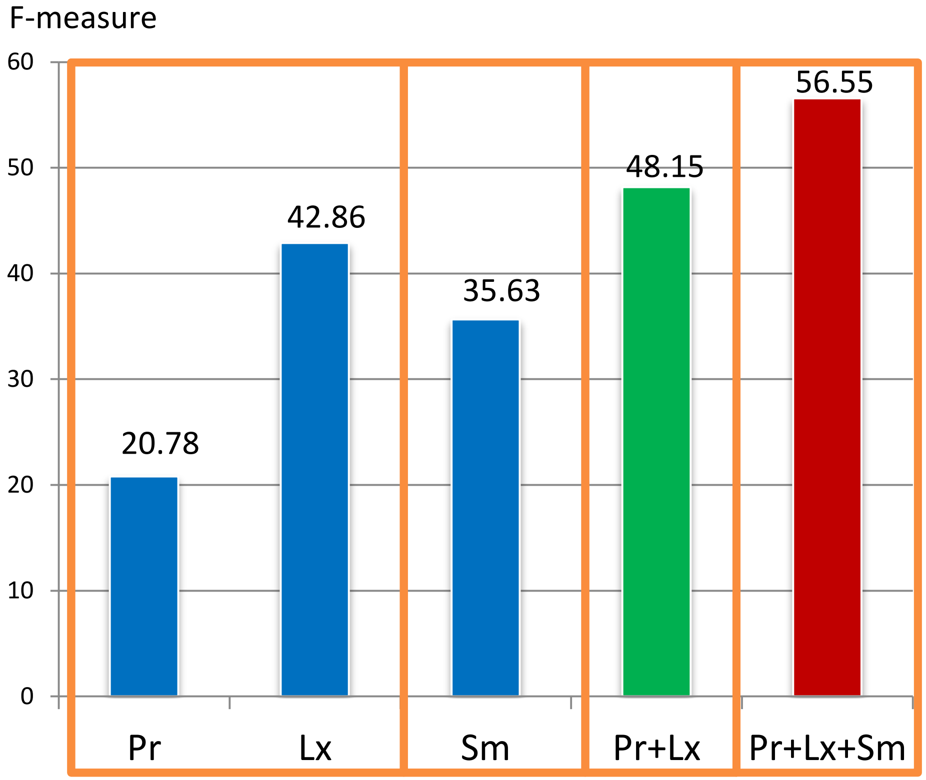
- Reference Key Terms
  - Annotations from 61 students who have taken the course
    - If the k-th annotator labeled N_k key terms, he gave each of them a score of $\frac{1}{N_k}$, but 0 to others
    - Rank the terms by the sum of all scores given by all annotators for each term
    - Choose the top $\overline{N}$ terms form the list ($\overline{N}$ is average N_k)
  - $\overline{N}$ = 154 key terms
    - 59 key phrases and 95 keywords

Experiments

- Evaluation
  - Unsupervised learning
    - Set the number of key terms to be $\overline{N}$
  - Supervised learning
    - 3-fold cross validation

Experiments

- Feature Effectiveness
  - Neural network for keywords from ASR transcriptions



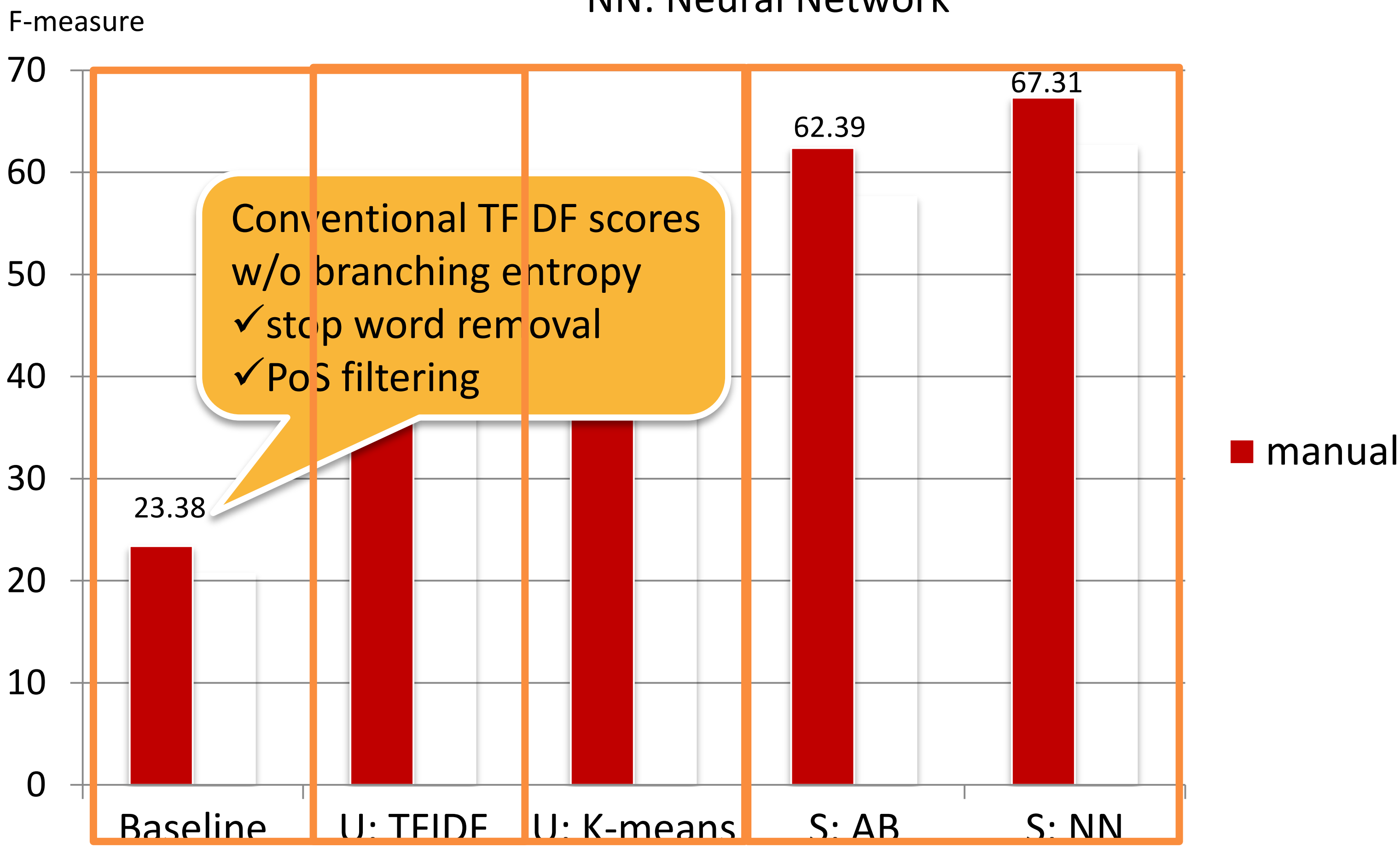
Pr: Prosodic
Lx: Lexical
Sm: Semantic

Three sets of features are all useful

Experiments

- Overall Performance

AB: AdaBoost
NN: Neural Network

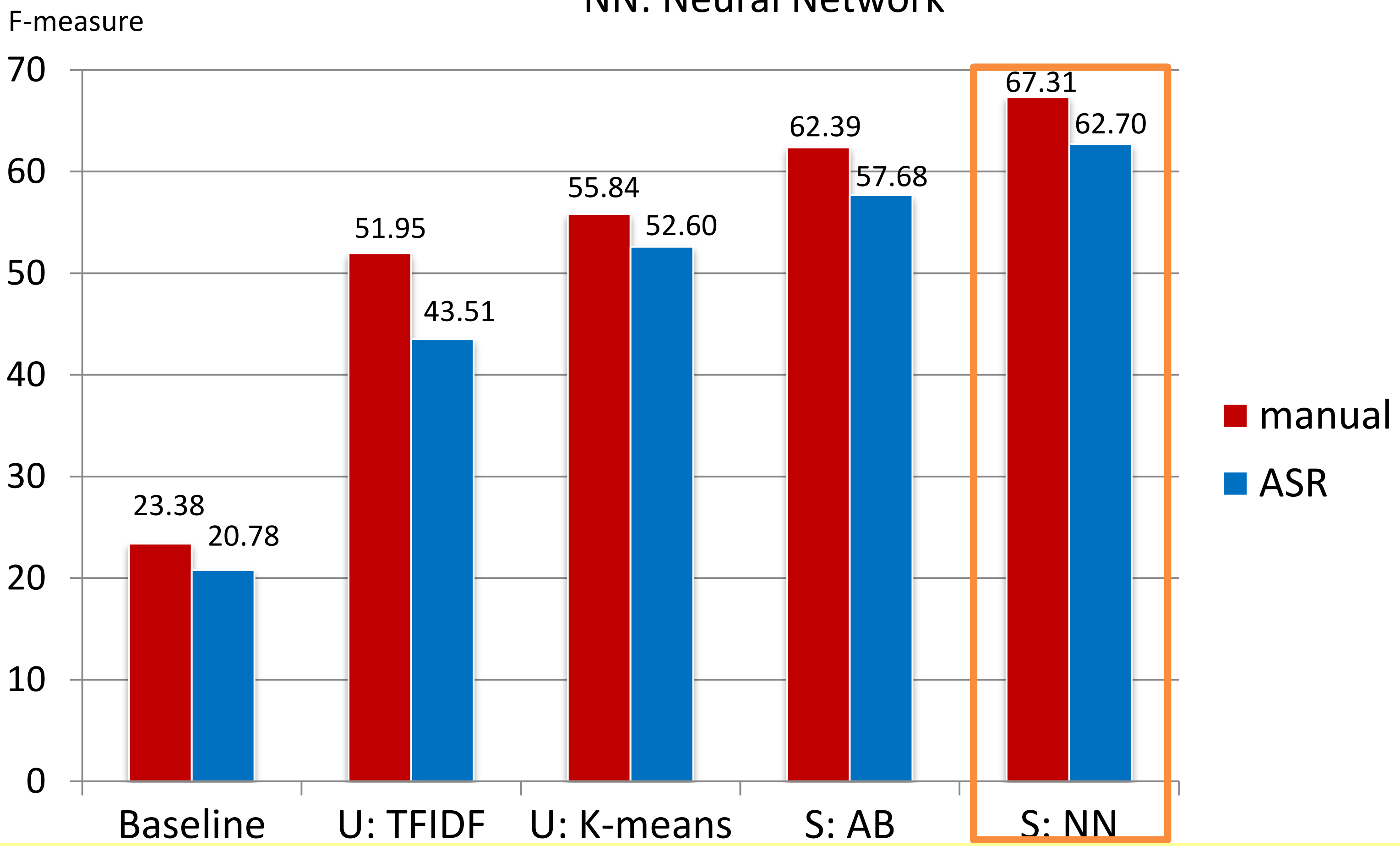


Supervised approaches are better than unsupervised approaches

Experiments

- Overall Performance

AB: AdaBoost
NN: Neural Network



Supervised learning using neural network gives the best results

Conclusion

Conclusion

- We propose the new approach to extract key terms
- The performance can be improved by
  - Identifying phrases by branching entropy
  - Prosodic, lexical, and semantic features together
- The results are encouraging

Thanks for your attention! ☺
Q & A

Thank reviewers for valuable comments

NTU Virtual Instructor: http://speech.ee.ntu.edu.tw/~RA/lecture