Abstract—This paper presents a two-stage approach to story segmentation and topic classification of broadcast news. The two-stage paradigm adopts a decision tree and a maximum entropy model to identify the potential story boundaries in the broadcast news within a sliding window. The problem for story segmentation is thus transformed to the determination of a boundary position sequence from the potential boundary regions. A genetic algorithm is then applied to determine the chromosome, which corresponds to the final boundary position sequence. A topic-based segmental model is proposed to define the fitness function applied in the genetic algorithm. The syllable- and word-based story segmentation schemes are adopted to evaluate the proposed approach. Experimental results indicate that a miss probability of 0.1587 and a false alarm probability of 0.0859 are achieved for story segmentation on the collected broadcast news corpus. On the TDT-3 Mandarin audio corpus, a miss probability of 0.1232 and a false alarm probability of 0.1298 are achieved. Moreover, an outside classification accuracy of 74.55% is obtained for topic classification on the collected broadcast news, while an inside classification accuracy of 88.82% is achieved on the TDT-2 Mandarin audio corpus.

Index Terms—Genetic algorithm (GA), segmental model, story segmentation, topic classification.

I. INTRODUCTION

As the number of audio/speech/video stream data increases dramatically, it is urgent to have an intelligent agent to automatically detect story boundaries and topics on audio/speech/video streams for further applications. For instance, a well-segmented and classified broadcast news stream is clearly a significant prerequisite for information retrieval, data mining, summarization, and management. Moreover, an explicit story document is more suitable than an entire audio document without story boundaries when browsing or retrieving a multimedia document.

The story segmentation problem discussed in this study can be simply defined as: given a stream of text articles or text transcriptions of a broadcast audio/speech/video stream via a speech recognizer, identify the locations where the event or topic changes. Previous studies have presented the approaches to story segmentation in a feature space or a model space. For feature-space approaches, lexical, audio-visual or prosodic information is generally employed. Wong et al. [6] adopted lexical features to capture words or phrases that are commonly used at the beginning or end of a segment in a particular domain. Moreover, C99 [7] and TextTiling [8] analyzed lexical cohesive strength and SeLect [9] utilized lexical chaining between textual units for text segmentation. In contrast to the approach based on text features, Hauptmann et al. [1] integrated closed-caption cues with the results of image and audio processing, while restricting content analysis to some specific TV programs. Shriberg et al. [32] and Tür et al. [33] presented a prosody-based method that obtained segmentation cues directly from the acoustic signal by pause, speaker change, pitch and energy. Second, the studies in the model space adopted topic transition models, exponential models, statistical models (hidden Markov models, HMMs) and hybrid models. Makhoul et al. [2] presented the topic transition concept for story segmentation. The CMU approach [5] adopted content-based features, lexical features and an exponential model. Van Mulbregt et al. [3] presented an HMM-based scheme, which adopted a classical language model and a topic transition model for story segmentation. Utiyama et al. [26] presented a statistical model for domain-independent text segmentation. Furthermore, the Umass approach [4] adopted Local Context Analysis (LCA) methods, but some problems still remained. First, the approach is somewhat slow since it needs a database query per sentence for query expansion and concept retrieval. Second, for automatically recognizing the speech data without manually tagged sentence boundaries, the performance of the LCA concept retrieval might degrade if the sentence boundaries detected from ASR are inaccurate.


Although story segmentation and topic classification have been extensively studied, there still exist some problems. First, most story segmentation approaches adopt the sliding window
strategy for lexical, audio-visual or prosodic features. However, the sliding window strategy may incur local false alarms due to a small window size, or contrarily, miss the true boundaries for a large window size. Even if a well-tuned window size is given, the fixed-length window is still not robust for different types of data. Therefore, a global and efficient search strategy is required since the search complexity is enormous for story segmentation. Several search strategies have been proposed in previous research, such as hill climbing. Nevertheless, the aforementioned search strategies usually result in locally optimal solutions since the initial settings will affect the search result significantly. Genetic algorithms (GAs) [18]–[20] have been shown to be a good solution to search problems with high computational complexity in a multi-direction way. A GA can also avoid locally optimal solutions, due to the random initial chromosomes and the genetic operators: crossover and mutation. Since the story segmentation problem is a stream-type data segmentation problem and has a very large search space for the optimal solution, a GA is appropriate for story boundary detection. Furthermore, some previous studies solved the stream-type data segmentation problem for time series data and speech signals by evolutionary computation and GAs. Chung et al. [27] presented an evolutionary approach to pattern-based time series segmentation and obtained a satisfactory performance. Moreover, Salcedo-Sanz et al. [28] adopted GAs to solve the offline speaker segmentation problem for speech signals. Second, the HMM-based approach models the topic transition in broadcast news by state transition probability. However, unlike speech signals, topic transitions cannot be modeled easily, since they depend on the editorial style and the number and importance of the events, which vary daily. Therefore, story segmentation results by the HMM-based scheme are generally dominated by the observation probability of the HMM. However, topic information is useful for story segmentation since most stories represent significant topics in broadcast news. It is thus beneficial to incorporate topic information with lexical features to solve the story segmentation and topic classification problems simultaneously.

This study presents a topic-based segmental model by fusing the topic information with boundary word information. Since story segmentation is an NP-complete problem, GAs, a useful approach to NP-complete problems, are adopted in this study to overcome the aforementioned difficulties. When applying a GA to problem solving, a well-designed fitness function is critical. In this paper, a topic-based segmental model comprising a topic model and a boundary model is proposed to characterize the fitness function. The topic model uses the topic information to represent the coherence of topics inside a story. The boundary model uses the word information around the boundary to represent the distinction between two stories segmented by the boundary. Furthermore, in practice, the search complexity in story segmentation is $O(2^n)$ where $n$ denotes the number of words in the whole broadcast news program. A prefiltering procedure is beneficial to reduce the computational complexity. Therefore, a hybrid approach using a decision tree and maximum entropy is utilized for pre-filtering. In the two-stage story segmentation paradigm, stage 1 adopts the decision tree and maximum entropy model to obtain the coarse story boundaries locally using lexical features. This stage yields a very low miss probability but high false alarm probability, thus reducing the search complexity for further computation in the GA. In stage 2, the problem for story segmentation can be considered as an optimization problem, which can be handled by the GA. Story segmentation is thus transformed into the global determination of chromosomes (boundary position sequence) in the potential story boundary region based on a defined fitness function derived from a topic-based segmental model, which models the topic and boundary characteristics jointly in a segment. The topic information compensates for the insufficiency of the lexical information (boundary characteristics). Finally, the determination of the story boundaries (chromosomes), which maximize the fitness function (topic-based segmental model), is achieved globally by the GA.

The rest of this paper is organized as follows. Section II provides the system description. Section III defines the LSA-based Naïve Bayes topic classifier. Section IV gives a brief background introduction on GAs. Section V describes in detail the topic-based segmental model and genetic algorithm for story segmentation. Section VI presents experiments and the performance evaluation of the proposed method. Conclusions are given in Section VII.

II. SYSTEM DESCRIPTION

Fig. 1 shows the overall architecture for story segmentation and topic classification. In the training phase, a boundary model and a topic model are constructed from a large text corpus. The news text corpus with tagged topics was collected automatically from a news web portal every day. First, the features and the parameters of the boundary model for the decision tree and maximum entropy model were extracted and trained from the text corpus. Then a latent semantic analysis (LSA)-based Naïve Bayes topic classifier was constructed as the topic model for topic classification in the topic-based segmental model.

In the segmentation phase, given a speech document input, a speech recognizer was first adopted to transcribe the broadcast news into a top-5 syllable lattice as well as the potential word sequences. The syllable-based lattice or the word-based sequences was used for topic classification and segmentation. After extracting the features for pre-segmentation, the decision tree and maximum entropy model were adopted to obtain a coarse story boundary sequence using a sliding window. Finally, the genetic algorithm along with the topic-based segmental model was adopted to refine the story boundaries in speech documents based on the boundary model and topic model.

III. LSA-BASED NAÏVE BAYES TOPIC CLASSIFIER

Given a speech stream comprising several contiguous stories, a speech recognizer is adopted to transcribe the speech signal into a syllable or word sequence $X = (u_1, u_2, \ldots, u_k_1, \ldots, u_k_n, w_n)$, where $b$ denotes the known number of story boundaries, and $K = (k_1, k_2, \ldots, k_b)$ denotes the position sequence of known story boundaries. Given the context of a story $X_j = (u_k_{j-1}, u_k_{j-1+1}, \ldots, u_k_j)$,
the posterior probability \( p(C_q | X_j) \) for the word sequence \( X_j \) belonging to topic \( C_q \) is estimated based on Bayes’ theorem as

\[
p(C_q | X_j) = \frac{p(X_j | C_q)p(C_q)}{\sum_{m=1}^{Q} p(X_j | C_m)p(C_m)}
\]

(1)

where \( Q \) denotes the number of topics adopted in this study.

To formulate the topic model \( p(X_j | C_q) \), the famous concept of TF × IDF is exploited to obtain the topic information matrix, and LSA [16], [17] is then adopted to reduce the dimensionality of the feature space. First, a word-by-topic matrix is established from the news text training corpus. The topic information matrix \( M_{P \times Q}^{TM} \) is calculated as follows:

\[
M_{P \times Q}^{TM} = [t_{wq}]_{P \times Q}
\]

(2)

where \( P \) denotes the number of words, and \( Q \) denotes the number of topics. A text training corpus grouped into \( Q \) topics is utilized to generate the topic information matrix \( M_{P \times Q}^{TM} = [t_{wq}]_{P \times Q} \). Each element \( t_{wq} \) of the topic information matrix is defined as the similarity between word \( w \) and topic \( q \) estimated as follows:

\[
t_{wq} = \frac{tf_{wq}}{\sum_{i=1}^{P} tf_{iq}} \times (1 - E_w), 0 \leq t_{wq} \leq 1
\]

(3)

where \( tf_{wq} \) denotes the term frequency of word \( w \) in topic \( q \), and \( \sum_{i=1}^{P} tf_{iq} \) denotes the summation of the term frequencies of all the terms in topic \( q \). \( E_w \) denotes the entropy of word \( w \) over all topics:

\[
E_w = \frac{1}{\log(Q)} \sum_{i=1}^{Q} \left( \frac{tf_{wi}}{\sum_{m=1}^{Q} tf_{wm}} \right) \log \left( \frac{tf_{wi}}{\sum_{m=1}^{Q} tf_{wm}} \right)
\]

(4)

where \( \sum_{m=1}^{Q} tf_{wm} \) denotes the summation of the term frequency of term \( w \) over all topics, and \( 0 \leq E_w \leq 1 \). If \( E_w \) approaches 1, then word \( w \) has a low topic discriminability; otherwise, \( w \) has a high topic discriminability.

In addition, the topic information matrix \( M_{P \times Q}^{TM} \) may contain noise information, such as the words with very low topic discriminability. That is, the feature space contains redundant information. Therefore, LSA is introduced to reduce the dimensionality of the feature space by eliminating the noise information and extracting the latent semantic information between words and topics. LSA includes two functions, Singular value decomposition (SVD) and feature dimensionality reduction. The SVD projection is performed by decomposing the word-by-topic matrix \( M_{P \times Q}^{TM} \) into the product of three matrices, \( U_{P \times K}, S_{K \times K} \) and \( V_{K \times Q} \):

\[
M_{P \times Q}^{TM} = U_{P \times K} S_{K \times K} V_{K \times Q}, \text{ with } K = \min(P, Q)
\]

(5)

where \( U_{P \times K} \) and \( V_{K \times Q} \) have orthogonal columns, i.e., \( U^TU = V^TV = I \) and \( \text{rank}(M_{P \times Q}^{TM}) = \lambda \). \( S_{K \times K} = \text{diag}(\alpha_1, \alpha_2, \ldots, \alpha_K) \) denotes the diagonal matrix with \( \alpha_1 > 0, 0 \leq \alpha_i \leq \lambda \) and \( \alpha_i = 0, i > \lambda \). By reducing the ranks of the matrices \( U_{P \times K} \), \( S_{K \times K} \) and \( V_{K \times Q} \) to \( \lambda \), a truncated matrix \( \tilde{M}_{P \times Q}^{TM} \) is derived as follows:

\[
\tilde{M}_{P \times Q}^{TM} = U_{P \times \lambda} S_{\lambda \times \lambda} V_{\lambda \times Q}
\]

(6)

where \( \tilde{M}_{P \times Q}^{TM} \) denotes the least square approximation of \( M_{P \times Q}^{TM} \). The original feature space is transformed by multiplying \( U_{P \times \lambda}^T \) with \( \tilde{M}_{P \times Q}^{TM} \) to obtain a topic strength matrix \( M_{\lambda \times Q}^{SM} \) as follows:

\[
M_{\lambda \times Q}^{SM} = U_{P \times \lambda}^T \tilde{M}_{P \times Q}^{TM} = U_{P \times \lambda}^T U_{P \times \lambda} S_{\lambda \times \lambda} V_{\lambda \times Q} = S_{\lambda \times \lambda} V_{\lambda \times Q} = [t_{wq\lambda}]_{\lambda \times Q}.
\]

(7)

For topic classification, all the words in \( X_j = (w_{k_1j}, w_{k_2j}, \ldots, w_{k_lj}) \) are assumed to be independent. The word sequence \( X_j \) is first transformed into a vector \( T_{X_j} = [(t_{12}, t_{13}, \ldots, t_{lp})] \) by (3). The \( P \)-dimensional column vector \( T_{X_j \times p} \) is transformed into a \( \lambda \) dimensional row vector \( \tilde{T}_{X_j \times \lambda} \) by the following formula:

\[
\tilde{T}_{X_j \times \lambda} = T_{X_j} U_{P \times \lambda} S_{\lambda \times \lambda}.
\]

(8)

The vector space model (VSM), an algebraic model for representing text documents as vectors of index terms, is successfully used in relevancy rankings of documents. In this paper, based on the topic strength matrix \( M_{\lambda \times Q}^{SM} \), the likelihood of topic model \( p(X_j | C_q) \) between word sequence \( X_j \) and topic \( C_q \) is approxi-
mated using a vector space model (VSM) as follows:
\[
p(C_q|X_j) = \frac{\cos(T_{X_j}^T, M_{qXQ}^T M_{q})}{\sum X \cos(T_{X}^T, M_{qXQ}^T M_{q})} = \frac{T_{X_j}^T S^2 M_{q}^T M_{q}}{\sum X \cos(T_{X}^T, M_{qXQ}^T M_{q}) (||T_{X_j}^T S|| \times ||M_{qXQ}^T M_{q}||)}
\]
where \(M_{qXQ}^T M_{q}\) denotes the \(q\)th column vector in \(M_{XQ}^T M_{q}\). Finally, the posterior probability \(p(C_q|X_j)\) is then derived by (1) and (9) as
\[
p(C_q|X_j) = \frac{\hat{p}(X_j|C_q) \cdot p(C_q)}{\sum_{m=1}^{Q} \hat{p}(X_j|C_m) \cdot p(C_m)}
\]
where \(M_{qXQ}^T M_{q}\) denotes the \(q\)th column vector in \(M_{XQ}^T M_{q}\). The topic of a word sequence is determined by the following equation:
\[
C_q^* = \arg \max_{1 \leq q \leq Q} \hat{p}(C_q|X_j).
\]

IV. BACKGROUND TO GENETIC ALGORITHMS

Genetic algorithms (GAs) [18]–[20] are a technique which simulate genetic evolution to search for the solution to a problem, especially NP or NP-complete problems with large search complexity and high computational complexity. The story segmentation problem in this study is an NP-complete problem with \(O(2^n)\) computational complexity, thus the GA is suitable to search for a solution in this problem. GA uses multidirectional search to reduce search complexity from \(O(2^n)\) to \(O(n \times ps \times gn)\), where \(n\) denotes the number of words of the input word sequence, \(ps\) denotes the population size and \(gn\) denotes the generation number. Although there are the numerous optimization techniques, such as hill climbing and simulated annealing, the hill climbing algorithm will easily incur a locally optimal solution affected by the initial setting. In addition, the simulated annealing algorithm utilizes the Gibbs and Boltzmann distribution to avoid the local optima problem. Since the best solution with the largest temperature decrement is obtained by repeatedly search the same temperature region, the annealing algorithm has the problem of low efficiency. In contrast to the above two approaches, a GA can avoid incurring a locally optimal solution based on two special operators: crossover and mutation, which can search globally in a multidirection way among a large population.

Fig. 2 illustrates an example of the implementation of the GA for story segmentation. In a GA, an initial population, i.e., a set of chromosome, is created randomly. In this study, each chromosome maps to a possible story boundary sequence in each sub-segment which is generated by a rough story pre-segmenter. Then, two operators, crossover and mutation, are applied to those chromosomes based on two probabilities: crossover probability and mutation probability. The crossover operator interchanges two chromosomes with each other at a randomly selected position and the mutation operator changes one bit in a chromosome to its complement. Following these operators, the offspring are generated and evaluated by a pre-defined fitness function. The fitness function plays a critical role in GAs, defining the fitness a chromosome has. A roulette wheel is then created based on the fitness of the offspring, and better chromosomes are selected from the offspring by the roulette wheel. The roulette wheel has the function that chromosomes with higher fitness values are more likely to be selected for the next generation. Nevertheless, the best parent populations are kept for the next generation. The above procedure continues until the termination condition is satisfied, such as when a certain number of generations are produced or the difference of average fitness values in the population between two generations is below a threshold. Finally, the chromosome with the highest fitness value is regarded as the best solution for the problem.

V. STORY SEGMENTATION USING GENETIC ALGORITHM

In this study, the segmentation problem is transformed into an evolution problem, and then solved by the genetic algorithm. The story segmentation problem can be modified easily to fit a GA by mapping the story boundary positions to the chromosomes, and defining a fitness function to evaluate the possible story boundary position (the chromosome).

Fig. 3 shows the block diagram for story segmentation by GA. The input sequence is the form of a syllable or word sequence generated from the speech recognizer. This study adopts a decision tree and a maximum entropy model over the word sequence, and determines the potential boundaries locally. The potential boundary is then relaxed to include its neighborhood for further processing. The story segmentation problem is transformed into discovering a chromosome (boundary position sequence) in the potential boundary regions using the genetic algorithm shown in Fig. 2. Fig. 4 shows an example for problem mapping. The input word sequence is pre-segmented with three potential boundaries. A potential region is expanded to cover ten words centered on its corresponding potential boundary. Therefore, a possible segmentation solution can be presented as
A. Stage 1: Coarse Boundary Detection by Decision Tree and Maximum Entropy

This study adopts the decision tree [24] and the maximum entropy model [5], [24] to search the coarse story boundary locally within a sliding window. The aim of this stage is to reduce the search complexity of the GA under a very low miss probability but high false alarm probability. The speech recognizer transcribes the speech into syllable and word sequences. For the syllable/word sequence, a sliding window

\[
WS_i = (w_{i-1} + 1, w_{i-1} + 2, \ldots, w_{i-1} + \ell_{dw}, w_{i-1} + 1, w_{i-1} + 2, \ldots, w_{i-1} + \ell_{dw} - 1, w_{i-1} + \ell_{dw})
\]

with a length of \(2 \cdot \ell_{dw} + 1\) is defined to cover \(2 \cdot \ell_{dw} + 1\) words and shifts with a step size of \(\ell_{wb}\) in the segmentation procedure. The posterior probability \(p(SB|WS_i)\) is defined based on Bayes’ theorem

\[
p(SB|WS_i) = \frac{p(SB_i|SB) \cdot p(SB)}{\sum_{SB'} p(SB_i|SB) \cdot p(SB')}
\]

where the random variable \(SB \in \{YES, NO\}\) denotes whether \(ws_i\) is a story boundary, and \(p(WS_i|SB)\) denotes the boundary model. The prior probability \(p(SB)\) is assumed to be uniform, and is set to \(1/(2 \cdot \ell_{w} + 1)\).

Story boundary modeling requires a huge amount of training data to avoid the data sparseness problem. The boundary model \(p(WS_i|SB)\) is approximated by combining the scores from the decision tree \(p_{DT}\) and the maximum entropy model \(p_{ME}\) in the model domain

\[
p(WS_i|SB) = \beta \cdot p_{DT}(WS_i|SB) + (1 - \beta)p_{ME}(WS_i|SB)
\]

where \(\beta\) denotes a fusion factor. The segmentation procedure in Stage 1 was performed by Franz’ method [24]. Franz’ method is a hybrid approach to story segmentation using a fusing decision tree and a maximum entropy model. In the training procedure, training data is used to train a decision tree based on a set of binary features (the question set) and the final score \(p_{DT}\) to decide if \(ws_i\) is a boundary estimated at each leaf node. In another words, the training data is also used to train the model parameters of the maximum entropy model comprising a log-linear model based on a set of binary features. Therefore, the score \(p_{ME}\) of the maximum entropy model can be calculated by the feature set of input data based on the trained model parameters in the test phase. Similar to Franz’ method, the boundary model adopts three features in the classification and regression tree (CART).

1) The duration of events (silence or pause) marked as non-speech in the ASR transcript.
2) The presence of words and word pairs strongly correlated with the document boundaries.
3) The comparative distributions of nouns on either side of the proposed boundary \(w_i\).

Second, the maximum entropy model adopts three features.
1) The three features adopted by the decision tree model.
2) The n-grams obtained from the context in \(W_i\) to the left and right of the proposed boundary \(w_i\).
3) Structural and large-scale regularities in the broadcast news, including specific time slots for commercials and the speaking rate.

In the segmentation procedure, the posterior probability \( p(SB|WS_k) \) of each sliding window \( WS_k \) is estimated and depicted as a score curve across the entire text stream. A peak extraction procedure is adopted to extract the coarse story boundaries, defined by a posterior probability higher than a pre-tuned threshold \( \varepsilon \). The pre-tuned threshold epsilon was tuned to achieve a target miss rate in the coarse boundary detection of less than 0.01 based on a tuning data set.

B. Stage Two: Fine Boundary Detection

1) Topic-Based Segmental Model: Since a GA is adopted for fine boundary detection, a well-designed fitness function is needed to formulate the segmentation problem. The topic-based segmental model is employed to incorporate the topic and boundary word information in characterizing the fitness function for global formulation of story segmentation. Therefore, the GA can work with the topic-based segmental model for fine boundary detection based on the detected coarse story boundaries in Stage 1.

Given a word sequence \( X = (w_1, w_2, \ldots, w_n) \) comprising several stories with unknown story boundaries, \( b \) denotes the number of story boundaries, \( K = (k_1, k_2, \ldots, k_b) \) denotes the story boundary position sequence and \( C = (C_1, C_2, \ldots, C_{b+1}) \) denotes the topic index sequence corresponding to the stories. Assuming \((b, K)\) and \( C \) are conditionally independent given \( X \), for a given word sequence \( X = (w_1, w_2, \ldots, w_{k_1}, \ldots, w_{k_b}, \ldots, w_n) \) with story boundary number \( b \), story boundary position sequence \( K \) and topic index sequence \( C \), the posterior probability \( p(b, K, C|X) \) is estimated as follows by the topic-based segmental model based on the Bayes’ theorem

\[
p(b, K, C|X) = \frac{p(b)p(K)p(C)p(X|K,b)p(X|C)p(X|b)}{Z}
\]

where the denominator \( Z \) is a constant given the same word sequence \( X \) in the story segmentation procedure. In this framework, the most important components are the boundary model \( p(X|K,b) \) and the topic model \( p(X|C) \).

First, assuming that all stories are conditionally independent given \((K,b)\), the boundary model \( p(X|K,b) \) can be derived from the boundary word information as

\[
p(X|K,b) = \prod_{j=1}^{b+1} p(X_j|k_{j-1},k_j)
\]

where \( X_j = (w_{k_{j-1}}, w_{k_{j-1}+1}, \ldots, w_{k_j}) \) denotes the word sequence in story \( j \), and \( w_{k_0} \) and \( w_{k_{b+1}} \) correspond to \( w_1 \) and \( w_n \), respectively. Additionally, this study makes a geometric mean approximation for \( p(X_j|k_{j-1},k_j) \) by the boundary models for boundaries \( w_{k_{j-1}} \) and \( w_{k_j} \) based on (15) as

\[
\hat{p}(X_j|k_{j-1},k_j) \approx p(WS_{k_{j-1}}|SB)^\eta \cdot p(WS_{k_j}|SB)^{1-\eta}
\]

where \( \eta \) denotes a balancing factor to manage the effect of the start and the end boundary model, and is tuned based on the development data. In (16), the front preceding story boundary \( k_{j-1} \) and the succeeding story boundary \( k_j \) occur consecutively. The probability \( p(X_j|k_{j-1},k_j) \) can be approximated by the two independent posterior probabilities \( p(WS_{k_{j-1}}|SB) \) and \( p(WS_{k_j}|SB) \) from (13) using the geometric mean approximation.

Second, the topic model is defined as \( p(X|C) = \prod_{j=1}^{b+1} p(X_j|C) \), where \( p(X_j|C) \) can be formulated the same as in (9) under the assumption that all the stories are conditionally independent given \( C \). In order to control the effect of the topic model and boundary model, a geometric mean fusion is applied on the two models in (14) as follows:

\[
\tilde{p}(b, K, C|X) \propto \prod_{j=1}^{b+1} p(X_j|C)^\gamma \cdot \hat{p}(X_j|k_{j-1}, k_j)^{1-\gamma} \cdot p(K|b)p(b)p(C)
\]

where the balancing factor \( \gamma \) is adopted to balance the effect between the topic and boundary models, and is tuned based on the development data. In (17), the geometric mean approximation is used for fusing the contribution of the topic model and boundary model to deal with the different characteristics of different data/corpora.

In addition, the conditional probability \( p(K|b) \) in (17) denotes the prior probability of story boundary position sequence \( K \) associated with \( b \) story boundaries.

Because the word sequence can be considered as a sequence of binary random variables \( \{0,1\} \) where story boundaries occur, the probability \( p(K|b) \) of the \( b \) boundary positions sequence \( K \) in the word sequence can be characterized by a Poisson family distribution, i.e., \( p(K|b) = e^{-\theta b}/b! \). In this study, in order to reduce the number of model parameters in the topic-based segmental model, the probability \( p(K|b) \) is simplified and assumed to have a binomial distribution

\[
p(K|b) \approx \binom{n-1}{b} \theta^b (1-\theta)^{n-1-b}
\]
Finally, the posterior probability \( \tilde{p}(b, K|X) \) is expressed as follows:

\[
\tilde{p}(b, K|X) = \frac{1}{2n-1} Q^{b+1} 
\cdot \left( \frac{n-1}{b} \right)^{b+1} \prod_{j=1}^{b+1} \left[ p(X_j|C) \gamma \cdot \tilde{p}(X_j|k_{j-1}, k_{j})^{1-\gamma} \right]. \tag{19}
\]

In the segmentation paradigm, Stage 1 obtains the tentative number of story boundaries \( \tilde{b} \) and the tentative story boundary position sequence \( \tilde{K} \). This significantly reduces the search complexity in the GA. The GA and the topic-based segmental model \( \tilde{p}(\tilde{b}, \tilde{K}, \tilde{C}|X) \) are then adopted to search for a precise story boundary number \( b^* \) and the boundary position sequence \( K^* \) that maximize \( \tilde{p}(\tilde{b}, \tilde{K}, \tilde{C}|X) \).

2) Search Precise Boundary Positions Using Genetic Algorithm: All tentative topic boundary positions \( \tilde{K} = (\tilde{K}_1, \tilde{K}_2, \ldots, \tilde{K}_t) \) formed in Stage 1 are relaxed forward and backward to form \( b \) search regions \( S = (s_1, s_2, \ldots, s_t) \), where \( s_t = (u_{k-1, t-1}/2, u_{k-1, t-1}/2 + 1, \ldots, u_{k-1, t+1}) \) with a length of \( l_k \). A precise boundary in each search region is determined over the whole document by the genetic algorithm and the topic-based segmental model. In general, the GA adopts a sequence of numbers, generally called a “chromosome” or “individual”, to denote a solution set [18]. For problem mapping, this study transforms the local boundary position sequence within the search regions \( S = (s_1, s_2, \ldots, s_t) \) into a number sequence (chromosome): \( \text{chromosome} = (k_1, k_2, \ldots, k_t) \), where \( k_t \in [0, l_k + 1] \), and \( k_t = 0 \) denotes a lack of story boundaries inside the search region. Overall, each number in a chromosome denotes the boundary position inside the search region. For example, Fig. 4 shows an input word sequence with length \( n \) comprising several unknown story boundaries which are pre-segmented to generate three potential boundaries. Then a search region with a length of ten words is expanded for each potential boundary. Therefore, a story boundary solution can be represented as a chromosome. In this figure, the number with the three digits 379 represents that the input word sequence has three fine story boundaries which occur in the third word of the first search region, the seventh word of the second search region and the ninth word of the third search region, respectively.

In the GA, the design of the fitness function plays the most important role in discovering the solution. Based on the topic-based segmental model, the fitness function is defined as follows:

\[
\text{Fit(chromosome)} = \max_{C} \tilde{p}(b, K, C|X) = \max_{C} \frac{1}{2n-1} Q^{b+1} 
\cdot \left( \frac{n-1}{b} \right)^{b+1} \prod_{j=1}^{b+1} \left[ p(X_j|C) \gamma \cdot \tilde{p}(X_j|k_{j-1}^u, k_{j}^u)^{1-\gamma} \right]. \tag{20}
\]

where \( \text{chromosome}^u = (k_1^u, k_2^u, \ldots, k_t^u) \), and \( k_t^u \in [0, l_k + 1] \) denotes the \( u \)th chromosome or solution in the search region, and represents one case of boundary position sequence \( K \) bounded in the search regions.

In the evolution process of the GA, chromosomes with better fitness denote better solutions, and survive with higher probability from generation to generation. The fine story boundaries with \( b^* \) boundaries and the position sequence \( K^* \) are obtained after the GA operations (crossover and mutation) and evaluation procedure shown in Fig. 2 are performed, and the termination condition is satisfied (fitness converges). Restated,

\[
(b^*, K^*) = \arg \max_{\text{chromosome}^u} \text{Fit(chromosome)}^u. \tag{21}
\]

VI. EXPERIMENTAL RESULTS

This section first presents the experimental materials and evaluation metrics, followed by the performance of the LSA-based Na"ive Bayes classifier. A parameter tuning procedure of the GA-based story segmenter is then described. Finally, a segmentation performance, \( t \)-test, an execution time and training time comparison between the proposed approach and Makhoul’s [2], Franz’s [24] and HMM-based [3] approaches are demonstrated on a test set based on the tuned parameters.

A. Training, Development, and Test Set

A news text corpus was collected for topic classifier training. The news text corpus was gathered daily from the United Daily News website (UDN) [21] over a period of four months (March–June of 2002). The UDN Corporation is one of the biggest publishers of print media in Taiwan, and the news catalog on the UDN News website is almost complete. This corpus comprises 46,965 stories or 23 million words with 14 main topics (e.g., Domestic, World, Politics, Stock, Business, IT, Health, Science, Travel, Entertainment) and about 110 subtopics (each main topic has on average eight subtopics). This study adopted the 110 subtopics as the topic classification catalog.

Two audio corpora were adopted for story segmentation evaluation: the BCC (Broadcasting Corporation of China) corpus and the TDT (Topic Detection and Tracking) Mandarin corpus. Because the TDT corpus has the characteristic that contiguous stories may belong to the same topic, the contribution of the topic model to the story segmenter may decrease. The BCC corpus was utilized to perform an individual experiment to evaluate the effect of the topic model on the story segmenter, since in that corpus, contiguous stories belong to different topics. The BCC audio corpus was collected from the Broadcasting Corporation of China News radio channel in the corresponding period to the UDN corpora. The audio data were recorded and digitized at a sampling rate of 16 kHz with a 16-bit resolution. Twelve recordings (about 1 h of speech material) were manually transcribed and segmented into stories for tuning (six recordings) and testing (six recordings). Each recording includes five to seven short news stories produced by an anchor speaker, and each story contains about 10–30 sentences.

The TDT audio corpus was the Mandarin audio corpus from the Linguistic Data Consortium (LDC), named Topic Detection
and Tracking Phase 2 and Phase 3 (TDT-2 and TDT-3). The TDT-2 and TDT-3 Mandarin audio corpora were recorded from Voice of America (VOA) news broadcast. The TDT-2 VOA Mandarin audio data were collected during a five-month period (February—June of 1998), and comprised 2265 stories with 70 172 words.

TDT-3 VOA Mandarin audio data were collected during a three-month period (October—December of 1998), and comprised 3371 stories with 331 494 words. The TDT-2 corpus comprised 100 topics defined in LDC, and only 564 stories corresponding to those topics. Similarly, TDT-3 contained 100 topics, which were different from those in TDT-2, and only 1542 stories corresponded to the topics. Word transcriptions from the Dragon Mandarin speech recognizer were adopted in this experiment. Because the transcription result of the TDT corpus does not include syllable lattice data, the syllable-based approach was not applied to it. The TDT-2 (564 stories) and one-third of the TDT-3 corpus (542 stories) were adopted as the training set for topic classification. One-third of the TDT-3 corpus (500 stories) was adopted as the development set for story segmentation, and the remaining part of the TDT-3 corpus (500 stories) was adopted as the test set for story segmentation. Additionally, the $t$-test was performed for significance testing by randomly dividing the test set into ten subsets.

In practice, since the broadcast news audio data include many foreground and background events, the audio data have to be segmented and identified into different segments (e.g., anchor, report, commercial and noise) based on the acoustic features first using an audio segmentation and classification system [30]. This study adopted only anchor segments for the potential story stream, because the word error rate in the transcription of reporter speech is too high. However, an anchor segment include more than one story, and must be further segmented and identified by story segmentation and topic classification approaches based on the context of the segment. Only the BCC corpora were processed using the audio segmentation and classification system. Because the TDT corpora have fewer speaker or background noise changes than the BCC corpora, and the Dragon system was adopted to process the audio segments, such as the music in the beginning of a show. The TDT corpora were not processed using the audio segmentation and classification but the standard word transcription was adopted for evaluation.

B. Performance Evaluation of Story Segmentation

In this experiment, story segmentation accuracy was evaluated by the standard TDT metric with a 15-s tolerance window [23], [31]. The miss probability $p_{miss}$ and false alarm probability $p_{fa}$ were considered. Moreover, two joint costs $C_{seg}$ and $C_{norm}$ were also adopted for evaluation:

$$C_{seg} = C_{miss} \cdot p_{miss} \cdot p_{target} + C_{FA} \cdot p_{fa} \cdot p_{naive}$$

$$C_{norm} = \frac{C_{seg}}{\min(C_{miss} \cdot p_{target}, C_{FA} \cdot p_{naive})}$$

(22)

(23)

where $C_{miss}$ and $C_{FA}$ denote the costs of a miss and of a false alarm, respectively; $p_{target}$ denotes the a priori target probability; $p_{naive} = 1 - p_{target}$, and $C_{seg}$ denotes the bottom-line representation of the performance of a story segmentation task. However, the value of $C_{seg}$ is also a function of the parameters $C_{miss}$, $C_{FA}$, and $p_{target}$, and $C_{seg}$ is normalized so that $C_{norm}$ can be no less than 1 without extracting information from the source data. In this experiment, the parameter settings were the same as those of Doddington [31], i.e., $C_{miss} = 1, C_{FA} = 1, p_{target} = 0.3$ for the TDT-2 corpus, and $C_{miss} = 1, C_{FA} = 0.3, p_{target} = 0.3$ for the TDT-3 and BCC corpora. Furthermore, the execution and training time for each recording was also used to compare the computational complexity for each approach.

C. Audio News Transcription Using Speech Recognizer

For the BCC audio broadcast news, an HMM-based speaker-independent large vocabulary continuous speech recognizer with two-pass search strategy was adopted [22], [29] for broadcast news content transcription. The recognition models were trained by the spoken documents from Voice of Taipei (VOT), UFO station (UFO), Police Radio Station (PRS), and Voice of Han (VOH) from December 1998 to July 1999 [25]. The 5465 spoken documents (about 14 h of speech data) were divided into two parts, 4/5 for training and the rest for testing. The error rates for syllable-level and character level were 22% and 25%, respectively. For the TDT-2/TDT-3 audio corpus, the word transcription result from the Dragon Mandarin speech recognizer was adopted in the experiment. For the TDT-2 corpus, the Dragon speech recognizer achieved error rates of 35.38% for word level, 17.69% for character level and 13.00% for syllable level; For the TDT-3 corpus, the error rates were 36.97% for word level, 19.78 for character level and 15.06% for syllable level.

Two approaches, syllable-based and word-based, were adopted to construct the content information from the speech recognizer for story segmentation. The aim of the syllable-based approach is to employ more syllable features for story segmentation to overcome the high word error rate of speech recognition. The syllable segments range from 1-syllable to 4-syllable segments. For instance, the syllable sequence “cheng kung da hsueh (National Cheng Kung University)” can be converted into four syllable segments: “cheng,” “cheng kung,” “cheng kung da,” and “cheng kung da hsueh.” The top-five syllable sequences with the highest recognition scores were selected to generate the potential syllable segments for topic classification. The syllable-based method converts the speech signal into base syllable segments from the top-five syllable sequences. Similarly, a word-based approach was adopted to convert the syllable sequence from the speech recognizer into a word sequence. The bi-gram language model was adopted to obtain the best five sentences as well as their corresponding potential word segments in order to include more possible features from top five best sentences to tolerate the ASR errors for topic classification. However, in the TDT-2/TDT-3 corpus, the transcription of the Dragon speech recognizer only comprises the top-one word sequence. The top-one word sequence was employed for the word-based experiment and a syllable sequence was transformed from the top-one word sequence directly to perform the syllable-based experiment.
D. Experiments on Topic Classification

In this study, the topic classifier plays a very important role in story segmentation. Therefore, topic classification evaluation was performed with the newswire news. In the news text corpus, 40,255 stories were adopted for training, and the remaining 6,710 stories were adopted for testing.

The aim of this experiment is to show the temporal effect of the topic classifier, which was trained by the static and dynamic models. The dynamic model trained on the most recent one week of collected news text was constructed to handle the short life cycle of the broadcast news, and the static model was trained on all the data. Fig. 5 shows the percentage of the number of days with continued news subtopics for the entire news text corpus during a period of four months. The figure indicates that about 90% of news events occurred in the most recent one week. Nevertheless, the temporal effect was not adopted in the TDT-2/TDT-3 topic classifier, since this phenomenon may not be significant in the VOA material.

In the training phase, a topic classifier was trained by the corpus collected in the most recent week. Additionally, another text corpus was collected to train the static model with all topics in our corpus. Table I shows the comparison of the classification accuracy of the dynamic and static models. In the test phase, the topic classifier accepted the input, and output a topic based on the dynamic model. For the inside test over the 40,255 training news stories, the system achieved a topic classification accuracy of 84.89%. For the outside test over 6,710 news stories, the system achieved 74.55% topic classification accuracy. By comparison, the static model achieved precision rates of 66.79% for the inside test, and 58.15% for the outside test. Overall, the dynamic model outperformed the static model. The proposed topic classifier achieved a topic classification accuracy of 74.55%. The topic classification experiment was also applied to the TDT-2/TDT-3 Mandarin audio corpus. For the inside test over the training stories, the system achieved a topic classification accuracy of 88.82% and 80.3% for TDT-2 and TDT-3, respectively.

E. Parameter Tuning on Story Segmentation

To obtain the best parameter setting for performance evaluation, a tuning procedure was constructed on the development data set. Two parameter sets, namely the pre-segmenter and GA-based story segmenter, were handled as follows.

1) Parameter Tuning of the DT_ME-Based Pre-Segmenter: The fusion factor \( \beta \) and the threshold \( \varepsilon \) of the hybrid pre-segmenter using the decision tree and maximum entropy were tuned first. The tuning range of \( \beta \) was from 0 to 1 with a step size of 0.25, and the tuning range of \( \varepsilon \) was from 0 to 1 with a step size of 0.1. Since the aim of the pre-segmenter is to minimize the miss probability first rather than the false alarm probability, the fusion factor \( \beta \) and threshold \( \varepsilon \) were set to 0.5 and 0.1, respectively, with a miss probability below 0.01 and with a false alarm probability of about 0.6 for both the BCC and TDT corpora.

2) Parameter Tuning of Window Size \( 2 \cdot I_{\text{win}} + 1 \): The window size \( 2 \cdot I_{\text{win}} + 1 \) was set to be half the average story length in previous tuning. In this subsection, the window size was tuned with a tuning range of \( 1/16, 1/8, 1/4, 1/2 \) and 1 average story length based on the optimized \( \beta \) and \( \varepsilon \). The tuning result indicates that the best performance was obtained by a window size of half of the average story length. Conversely, the performance degraded when the window size was set to the average story length. The reason is that the miss probability rose significantly since the large window size will miss many story boundaries.

3) Parameter Tuning of the GA-Based Story Segmenter: Since the GA-based segmenter had many parameters, the tuning procedure was split into two subprocedures. Because the parameters in the GA mainly affect the convergence speed when the population and generation number are large, the first subprocedure fixed the parameters in the GA empirically, with population size \( N_{\text{pop}} = 200 \), generation number \( N_{g} = 2000 \), crossover probability \( P_{c} = 0.5 \), and mutation probability \( P_{m} = 0.5 \). The tuning ranges of the balancing factors \( \gamma \) and \( \eta \) were then set from 0 to 1 with a step size of 0.2. Figs. 6 and 7 show the tuning results. For the BCC corpus, the best settings of the balancing factors \( \gamma \) and \( \eta \) were found to be 0.6. For the TDT corpus, the best settings of the balancing factors \( \gamma \) and \( \eta \) were 0.2 and 0.6, respectively.

In the second subprocedure, the balancing factors \( \gamma \) and \( \eta \) were fixed at the best settings, and the parameters in the GA were tuned based on the development data set. The tuning ranges of the parameters in GA were set as follows.

1) Population size \( N_{\text{pop}} \) from 20 to 100 with a step size of 20
2) Generation number \( N_{g} \) from 100 to 1000 with a step size of 100
3) Crossover and mutation probability \( P_{c}, P_{m} \); from 0 to 1 with a step size of 0.25, and \( P_{c} + P_{m} = 1 \).

---

**Table I: Topic Classification Accuracy by Dynamic Model and Static Model**

<table>
<thead>
<tr>
<th>Test Type</th>
<th>Dynamic Model</th>
<th>Static Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inside Test</td>
<td>84.89%</td>
<td>66.79%</td>
</tr>
<tr>
<td>Outside Test</td>
<td>74.55%</td>
<td>58.15%</td>
</tr>
</tbody>
</table>
the best-tuned parameter settings, respectively. The execution time evaluation was performed on a Pentium 4 1.5-GHz PC, and the time was measured in minutes.

1) Experiments on News Text Corpus: To demonstrate the upper bound of the GA-based story segmenter performance, the text corpus was tested for story segmentation. News text streams comprising several news stories in concatenation is generally not easy to collect. Therefore, a number of news stories were concatenated together randomly to form a long text as the testing document. The document comprising 3500 long texts obtained by concatenating five news stories randomly chosen from the UDN news Text. The evaluation results of the test corpus of the 3500 long texts using the same tuned parameters as the BCC corpus indicate a miss probability of 0.1197 and a false alarm probability of 0.1228. The result from the simulated data may be treated as a reference for the upper bound since all the data contain no speech recognition errors.

2) Experiments on BCC Mandarin Broadcast News Corpus: The BCC broadcast news corpus was assessed by three methods: syllable-based story segmentation, word-based segmentation and Makhoul’s method [2]. Table II shows the experimental results from the BCC broadcast news. A miss probability of 0.1587 and a false alarm probability of 0.0859 were observed in word-based story segmentation. Additionally, a miss probability of 0.1495 and a false alarm probability of 0.0871 were achieved in syllable-based story segmentation. For comparison, Makhoul’s method produced a miss probability of 0.1671 and a false alarm probability of 0.1425. The proposed approach outperforms Makhoul’s method since the proposed approach can take care of the story boundaries of the entire program, rather than two contiguous stories, using the GA. The use of boundary word information and the GA can refine the story boundaries globally, so that the reduction in the false alarm probability is significant compared to Makhoul’s approach.

In summary, the performance of the syllable-based approach performed slightly better than the word-based approach. Because the syllable-based approach considers the 1–4-syllable segments from the top-five syllable lattices, the segmenter can capture more correct syllable information than the word-based approach with a high character error rate.

Second, although Makhoul’s method achieved a slightly lower miss probability than the others, it had a higher overall error probability than both of the word- and syllable-based approaches. The t value (5.98) exceeded the tabulated t value of 3.36 (p = 0.01), 8 degrees of freedom. This finding indicates that the difference between the proposed approach and Makhoul’s method is highly significant.

3) Experiments on TDT3 Mandarin Audio Corpora: In this section, only a word-based approach was experimented with on the TDT-3 Mandarin audio corpora, because the transcription units of the Dragon speech recognizer are words, not syllables. Table III shows the experimental results of the TDT-3 Mandarin corpus. The proposed approach achieved a miss probability of 0.1232 and a false alarm probability of 0.1298, and thus outperformed Franz’s approach and the HMM-based approaches [3] which involve a classical language model and a topic transition model for story segmentation. In particular, the reduction in the false alarm probability is significant because the GA
Fig. 8. Tuning results of the GA-based segmenter obtained by varying the population size $N_{pop}$, the generation number $N_g$, the crossover and the mutation probability $P_c$ and $P_m$.

<table>
<thead>
<tr>
<th>Table II</th>
<th>EXPERIMENTAL RESULTS ON BCC BROADCAST NEWS CORPUS (WITH $EXC_T$ GIVEN IN MINUTES AND $TRA_T$ GIVEN IN HOURS)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P_{miss}$</td>
</tr>
<tr>
<td>GA-based_W</td>
<td>0.1587</td>
</tr>
<tr>
<td>GA-based_S</td>
<td>0.1495</td>
</tr>
<tr>
<td>Makhoul’s</td>
<td>0.1671</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table III</th>
<th>EXPERIMENTAL RESULTS ON TDT-3 CORPUS (WITH $EXC_T$ GIVEN IN MINUTES AND $TRA_T$ GIVEN IN HOURS)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P_{miss}$</td>
</tr>
<tr>
<td>GA-based_W</td>
<td>0.1232</td>
</tr>
<tr>
<td>GA-based_S</td>
<td>0.1215</td>
</tr>
<tr>
<td>Franz’s</td>
<td>0.1293</td>
</tr>
<tr>
<td>HMM-based</td>
<td>0.1956</td>
</tr>
</tbody>
</table>

with the topic-based segmental model can obtain global consideration among the whole word stream. Nevertheless, the reduction in the miss probability is small since the contribution of the topic model is small; the boundary model has a higher weighting in the topic-based segmental model. The contribution of the topic model in the TDT corpus was found to be less than that in the BCC corpus. We believe that the broadcast news program flow of the VOA Mandarin corpus is different from that of the BCC corpus. VOA broadcast news is generally a report on a special topic in a whole one-hour program. Therefore, many contiguous stories have the same topic, and the boundary model plays a significant role. Additionally, a miss probability of 0.1215 and a false alarm probability of 0.1307 were achieved in syllable-based story segmentation. Although the syllable-based approach takes care of the high word error rate for story segmentation, the feature space is huge for story segmentation. In contrast to the word-based approach, the range of performance improvement of syllable-based approach for story segmentation is not as good as that in speech recognition. Furthermore, a GA-based approach, both word-based and syllable-based, needs more execution time than the others since word sequences are very long in the one-hour show.

VII. CONCLUSION

This study presented a two-stage approach to story segmentation and topic classification on Mandarin broadcast news using a topic-based segmental model and the genetic algorithm. In this approach, a speech recognizer is first adopted to generate the syllable- and word-based transcriptions from the audio broadcast news. Decision trees and maximum entropy are then adopted to identify the potential story boundaries locally. The genetic algorithm is then employed to obtain the final story
Experimental results indicate that the GA-based approach generalizes globally. Moreover, a topic-based segmental model outperforms the traditional methods for both syllable- and word-based segmentation approaches.

REFERENCES


Chia-Hsin Hsieh

Chia-Hsin Hsieh received the B.S. degree in electronics engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1981, and the M.S. and Ph.D. degrees in electrical engineering from National Cheng Kung University, Tainan, Taiwan, in 1987 and 1991, respectively.

Since August 1991, he has been with the Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan. He became a Professor and Distinguished Professor in August 1997 and August 2004, respectively. From 1999 to 2002, he served as the Chairman of the Department. He also worked at Massachusetts Institute of Technology Computer Science and Artificial Intelligence Laboratory, Cambridge, in summer 2003 as a Visiting Scientist. He was the Editor-in-Chief for International Journal of Computational Linguistics and Chinese Language Processing from 2004 to 2008. He serves as the Guest Editor of the ACM Transactions on Asian Language Information Processing, and the EURASIP Journal on Audio, Speech, and Music Processing in 2008–2009. He is currently in Editorial Advisory Board of The Open Artificial Intelligence Journal and the Subject Editor on Information Engineering of the Chinese Institute of Engineers. His research interests include speech recognition, text-to-speech, and spoken language processing.

Dr. Wu is a member of the International Speech Communication Association (ISCA), the Association for Computational Linguistics (ACL), and Association for Computational Linguistics and Chinese Language Processing (ACLCLP). He serves as a Guest Editor for the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING.