TOPIC FOCUSING MECHANISM FOR SPEECH RECOGNITION BASED ON PROBABILISTIC GRAMMAR AND TOPIC MARKOV MODEL

Takeshi Kawabata

NTT Basic Research Laboratories
3-1 Morinosato-Wakamiya, Atsugi-shi 243-01, JAPAN

ABSTRACT

This paper describes a new stochastic topic focusing mechanism for reducing the perplexity of natural spoken languages. In this mechanism, a predictive context-free grammar (CFG) parser analyzes input speech and generates grammar-rule sequences. These rule sequences drive a hidden Markov model (HMM), and the current topic is estimated as the HMM state distribution. The CFG rule probabilities are dynamically changed according to this topic state distribution. Evaluation of this mechanism using a large dialog text database confirms that it can effectively reduce the task perplexity.

1. INTRODUCTION

Speech conversation is the most comfortable interface for human-machine communication. For this purpose, we need large-vocabulary continuous speech recognition technologies that can cope with real-world speech. Since the accuracy of speech recognition systems tends to decrease as their vocabulary increases, it is necessary to apply focusing techniques to reduce the effective vocabulary size [1].

A great deal of research has been done on topic focusing mechanisms for speech recognition. These mechanisms work by identifying the current topic and restricting the number of candidate words accordingly. Kobayashi [2] used a topic tree, whose nodes correspond to restricted words, for keyword prediction. Yamashita [3] proposed a topic packet network (TPN) for generating a sentence template for the subsequent utterance. These methods assume that the current topic is always determined correctly. However, this assumption is often incorrect in real life, where incorrect recognition of a word can lead to fatal mis-identification of the entire topic.

This paper proposes a new stochastic topic focusing mechanism for reducing the perplexity of natural spoken languages. In this mechanism a finite-state hidden Markov model (HMM) is used as the topic transition model. The current topic is estimated as the state-probability distribution of this HMM. Because the system does not make any rigid decisions, fatal errors are completely avoided. For combining this stochastic topic transition model and grammar-based natural language processing, this paper uses a dynamic probabilistic grammar (DPG) framework [4]. The DPG is a context-free grammar (CFG) whose rule probabilities are dynamically controlled by a HMM. The grammar rule sequence, produced through syntactic processes, drives the topic Markov model and changes its state distribution. Each topic state has a probability table of CFG rules. By merging these tables according to the state distribution, the estimated topic is reflected in the CFG rule probabilities.

2. TOPIC FOCUSING MECHANISM BASED ON PROBABILISTIC GRAMMAR AND TOPIC MARKOV MODEL

2.1 Speech Recognition System with Topic Focusing

Figure 1 is a schematic diagram of a speech recognition/understanding system with the proposed topic focusing mechanism. The predictive CFG parser generates utterance candidates and drives phoneme verifiers. The parser searches for the candidate which achieves maximum probability by using the phoneme verification scores and a beam-search mechanism. The parser also generates a grammar-rule sequence while producing a candidate sentence. The rule

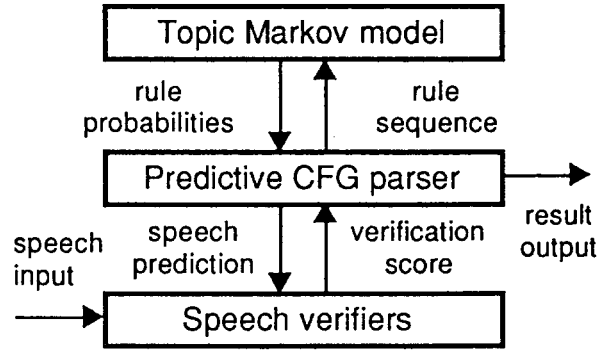

Fig. 1 Speech understanding system with stochastic topic focusing mechanism

0-7803-2431-5/95 $4.00

sequence drives the topic Markov model and changes its state distribution. The current topic is identified as this state distribution. Each topic state has a probability table for CFG rules. By merging these tables according to the state distribution, the estimated topic is reflected in the CFG rule probabilities.

2.2 Topic Transition Model

In this paper, the word "topic" refers to an information source used for predicting words and/or sentence structures. Grammar rules whose lhs (left hand symbol) is a pre-terminal symbol are called vocabulary rules. These rules expand a POS (part of speech) into a word. The other rules restrict the order of non-terminal symbols and compose a syntactic structure; these are called structure rules.

The topic Markov model is a finite-state ergodic HMM driven by grammar-rule sequences. Figure 2 shows this mechanism. The parser composes a sentence candidate and generates a grammar rule sequence. Here, the leftmost derivation method is used. The sequence of rule numbers drives the topic Markov model. Each state of the topic Markov model has a probability table for the grammar rules. Suppose that rule number k is sent to the topic Markov model. If state i has higher probability for rule k than the other states, the state distribution is concentrated on state i. In this way, our system estimates the current topic as the state-probability distribution. After that, the probabilities of all grammar rules are calculated for the next parsing step. Probability tables are merged according to the state-probability distribution. Usually, rules which belong to the same or similar topics have high probabilities in the same state. The system thus assigns higher probabilities to the grammar rules related to the estimated topic than to other rules. This focusing probability can be used for restricting the subsequent words and syntactic structures.

This focusing process continues through a dialog session. At the beginning of a dialog, the topic state probabilities are initialized uniformly. After processing an utterance, the topic state distribution is biased, and is inherited by the process for next utterance.

2.3 Mathematical Formulation

The state-probability distribution $\alpha(i,t)$ (of state i at step t) is calculated by the following procedure. At the beginning of a dialog, initialize the state probability at t=0 with a uniform value for each state.

$$\alpha(i,0) = 1/(number\ of\ states) \quad (1)$$

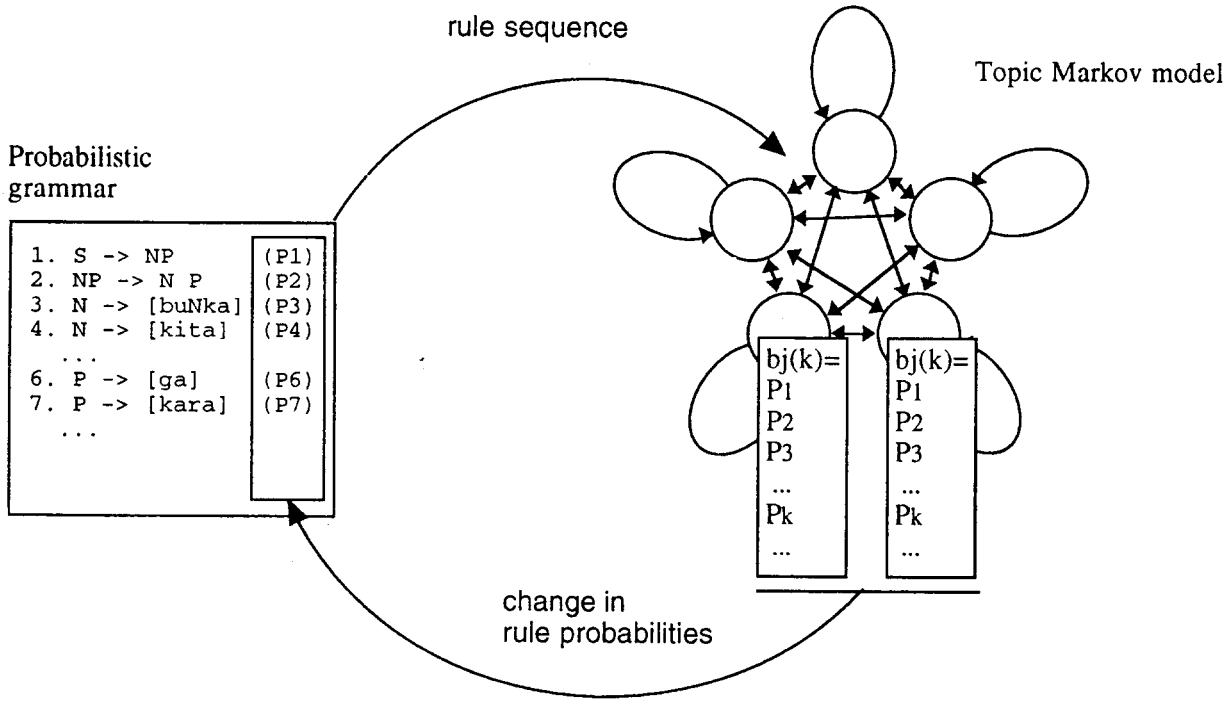

Fig. 2 Topic focusing mechanism based on probabilistic grammar and topic Markov model

On receiving rule number k, the HMM state-probability distribution changes according to the transition probabilities and the output probabilities.

$$\alpha(i,t) = \sum_j \alpha(j,t-1) a_{ji} b_i(k) \quad (2)$$

where

$\alpha(j,t)$: state-distribution probability (of state j at step t)

a_{ji} : transition probability (from state j to state i)

$b_i(k)$: output probability (to state i for rule k)

The production probability of grammar rule k for the next parsing step t is calculated as

$$P(k,t) = \frac{\sum_i \sum_j \alpha(j,t-1)\, a_{ji} b_i(k)}{\sum_l \sum_i \sum_j \alpha(j,t-1)\, a_{ji} b_i(l)} \quad (3)$$

where Σ_l is the sum over all rules containing the same left-hand symbol.

This normalization guarantees the consistency of the probabilistic grammar [5]. A probabilistic grammar G is consistent if

$$\sum_{x \in L(G)} p(x) = 1, \quad (4)$$

where $L(G)$ is the set of sentences generated from G.
Let V and W be the non-terminal and terminal symbol set of a given CFG. The probability summation for all sentences derived from a symbol X ($\in V \cup W$) is calculated as

$$P(X) = \sum_l p_l(R_1,\cdots,R_{t-1}) \prod_i P(Y_{li}), \quad (5)$$

where Y_{li} is the i-th rhs (right-hand symbol) of rule l, Π_i is the product over all rhs of rule l, and Σ_l is the sum over all rules containing the same lhs (left-hand symbol). Let $p_l(R_1,\cdots,R_{t-1})$ be the production probability of rule l dynamically changed according to the rule history. $P(Y_{li})$ can be calculated incrementally by replacing X of Equation (4) to Y_{li} and replacing $(R_1,\cdots,R_{t-1})$ to $(R_1,\cdots,R_{t-1},l)$. Finally, this recursive procedure arrives at terminal symbols. Here, define

$$P(X) \equiv 1 \quad (for\ X \in W), \quad (6)$$

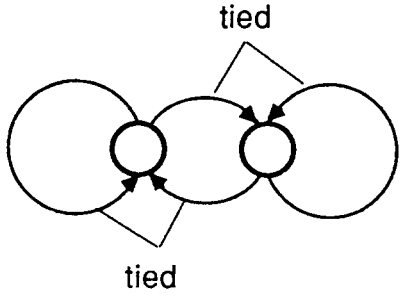

Fig. 3 Tied arcs of topic Markov model

and assume

$$\sum_l p_l(R_1,\cdots,R_{t-1}) = 1. \quad (7)$$

By tracing the above recursive procedure backwards,

$$P(X) = 1 \quad (X \in V \cup W) \quad (8)$$

for any symbol X. Consequently,

$$P(S) = 1 \quad (S : start\ symbol). \quad (9)$$

Thus, Equation (7) gives a sufficient condition for the consistency of a probabilistic grammar. The denominator of Equation (3) guarantees the consistency of our DPG framework. For the simple implementation of this normalization, the leftmost derivation parsing mechanism is required.

3. EVALUATION

3.1 Corpus

We evaluated the proposed topic focusing mechanism using a large dialog text database collected by simulating the dialog of a secretarial service at an international conference. In this simulation, two persons communicated through video display terminals. One played the role of a secretary, and the other played an applicant. The dialogs obtained in 354 sessions (over 23,000 sentences) were collected and transcribed into texts with POS labels. The vocabulary used in the dialogs amounted to 6,200 words. The first 300 dialog sessions are used for training; the other 54 sessions are used for tests.

3.2 Model Structure and Initialization

A topic Markov model is a finite-state ergodic HMM. Figure 3 shows an example of the ergodic structure. The model consists of several states and several arcs connecting the states. Each arc has a transition probability and symbol-output probabilities. In our current implementation, the arcs toward to the same state share the same output probabilities for reducing the degrees of freedom. These arcs are called tied arcs.

Generally, the convergence of HMM training is guided by the initial output probabilities. For setting them adequately, we categorize the given vocabulary into several word clusters according to their cooccurrence in a sentence using the BPD (Binomial Posteriori Distribution) word distance [6] and the LBG clustering method [7]. Taking the estimation reliability into account, adequate and robust word clustering can be done.

An HMM which contains the same number of states as the clusters is prepared. Each state corresponds to a word cluster. The initial output probability for a vocabulary rule (i.e. a word) is increased when the arc is tied to the states whose cluster contains the word. The other output probabilities are initialized with a lower value ($lower/higher \cong 10^{-6}$).

3.2 Training and Tests

The first 300 dialog sessions (20,000 sentences) were used to train the topic Markov model. The sentences in the training dialogs were parsed and transformed into grammar rule sequences with the leftmost derivation method, and each rule sequence from the beginning to the end of a dialog session was input to the topic Markov model. The forward-backward algorithm [8] was then used to train the topic Markov model.

We tested the perplexity reduction capability of the proposed topic focusing mechanism using the other 54 dialog sessions (3,000 sentences). There were no unknown words in this test data. Figure 4 shows the relationship between the number of HMM states (i.e. topics) and the word perplexity. As the number of states is increased, the perplexity is effectively reduced. Using 25 topic states, the effective vocabulary size (perplexity) is reduced to about 80. The perplexity reduction effects are summarized in Table 1. The vocabulary size of 6,200 is reduced by 80% by restricting the word order using a context-free grammar. Futhermore, the dynamic topic focusing mechanism reduces it by 98.7%.

CONCLUSION

A new stochastic topic focusing mechanism was proposed. Since our mechanism estimates the current topic as the state distribution of a topic Markov model (i.e. it does not make rigid decisions), fatal errors are completely avoided. Evaluation experiments using a large dialog text database confirmed that the proposed mechanism can effectively reduce the task perplexity.

REFERENCES

[1] Young, S. R., Hauptmann, A. G., Ward, W. H., Smith, E. T., and Werner, P. : "High Level Knowledge Sources in Usable Speech Recognition Systems," Comm. ACL, Vol.32, No.2, pp.183-194 (1989)

[2] Kobayashi, Y., Tanabe, M. and Niimi, Y.: "Keyword Prediction in a Speech Dialog System," Tech. Rep. JSAI, SIG-SLUD-9201-3, pp.19-26 (1992)

[3] Yamashita, Y. and Mizoguchi, R. : "Next Utterance Prediction Based on Two Kinds of Dialog Models," Proc. of Eurospeech '93, Berlin, pp.1161-1164 (1993)

[4] Kawabata, T. : "Dynamic Probabilistic Grammar for Spoken Language Disambiguation," ICSLP-94, S16-3.1, pp. 787-790 (Sep. 1994)

[5] Fu, K. S. : "Stochastic Languages for Picture Analysis," Computer graphics and image processing, Vol.2, pp.433-453 (1973)

[6] Tamoto, M. and Kawabata, T. : "Clustering Word Category based on Binomial Posteriori Cooccurrence Distribution," ICASSP-95, TP7 (May 1995)

[7] Linde, Y., Buzo, A., and Gray,R.: "An Algorithm for Vector Quantizer Design", IEEE Trans. Communication, COM-28, 1, pp.84-95 (1980)

[8] Levinson, S. E., Rabiner, L. R., and Sondhi, M. M.: "An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech recognition," BSTJ, Vol.62, No.4, pp.1035-1074 (1983)

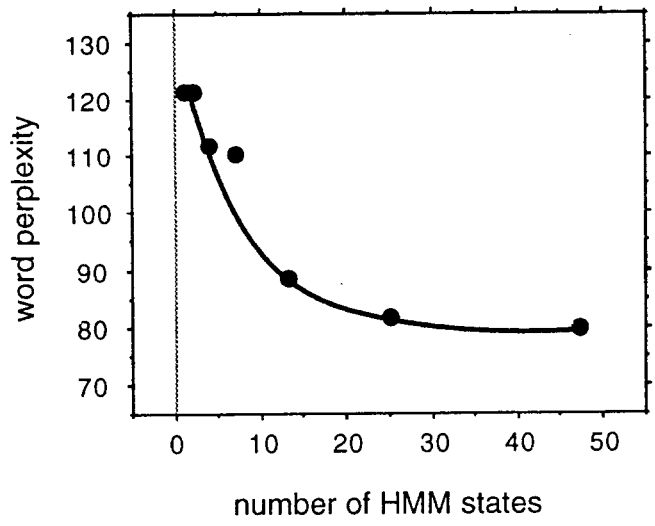

Fig. 4 Reduction of word perplexity

Table 1 Perplexity reduction by language models

language processing	word perplexity
none	6,200
context-free grammar	1,193
CFG + topic focusing (25 states)	80