

# Nested Joint Probability Model for Morphological Analysis and its Grid Pruning

**Koji FUJIMOTO, Nobuo INUI and Yoshiyuki KOTANI**

Tokyo University of Agriculture and Technology, Department of Computer Science  
2-24-16 Nakamachi, Koganei-city, Tokyo 184, Japan  
{kfujim,nobu,kotani}@cc.tuat.ac.jp

## Abstract

In recent work on morphological analysis based on statistical models, the conditional probability of the observed  $i$ -th word  $w_i$  with the  $i$ -th tag  $t_i$  after the  $(i-1)$ -th tag  $t_{i-1}$  is defined as the product of observation symbol probability and the state transition probability (i.e.  $P(w_i | t_i) \bullet P(t_i | t_{i-1})$ ). In order to improve accuracy, we face the following problems: 1) if we build hidden state levels using stricter categories (e.g. lowest POS class, over 3-gram, or word themselves), the state transition probability matrix becomes much bigger and more sparse; 2) if we use rough categories, the reliability of statistical information becomes lower in some parts of speech; and 3) the best state level is not the same among POS category, and some heuristic knowledge is necessary to select the best state structure.

This paper presents a novel stochastic model based on a nested tag class structure (Nested Joint Probability Model) (Fujimoto, Inui, and Kotani 1998) and its grid pruning technique for data resource saving. This model estimates morpheme transition probability by frequency lattice which is a direct product of two nested tag hierarchies. After an important frequency subset on the lattice is selected by using statistical test criterion for every pair of morphemes, the most reliable transition probability is estimated automatically without heuristic knowledge. Consider a word sequence  $A B$ . Class  $A_i$  ( $i=0, 1, \dots, m$ ;  $A_0$  means all) is a category in the  $i$ -th layer of the previous morpheme  $A (=A_m)$  in the word's hierarchy, and  $B_j$  ( $j=0, 1, \dots, n$ ) is in the  $j$ -th layer of the target morpheme  $B (=B_n)$ . The nested joint probability model is formulated as a series of joint probabilities,  $P(A_i B_j) = \{P(A_i B_{j-1}) \bullet P(A_{i-1} B_j) / P(A_{i-1} B_{j-1})\} \bullet I(A_i B_j)$  ( $i=0, 1, \dots, m$ ;  $j=0, 1, \dots, n$ ), where  $I(A_i B_j)$  is the conditional interaction effect of  $A_i$  and  $B_j$  when both  $A_{i-1}$  and  $B_{j-1}$  are observed. The frequency lattice of  $A$  and  $B$  is defined as the set of lattice  $F_{ij}$  that is the

joint frequency of  $A_i$  and  $B_j$  on a training dataset.

An illustration of frequency lattice is shown in Figure 1. For example, if a set of two frequencies  $\{F_{ih}, F_{kj}\}$  is selected through statistical test against the null

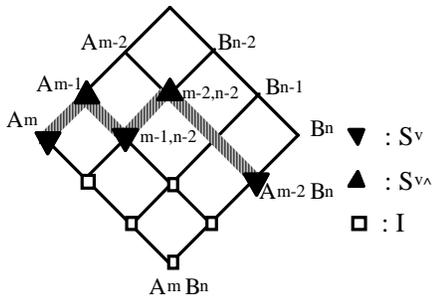


Figure 1 : Frequency Lattice

hypothesis of  $I(A_i B_j)=1$  corresponding to  $F_{ij}$  on the lattice, the conditional probability  $P(B/A)$  can be estimated by the optimized formula,  $F_{0n}/F_{0l} \bullet F_{il}/F_{ij} \bullet F_{kj}/F_{ko}$ .

The effectiveness of our model was tested, and compared with the established models: the Hidden Markov Model (HMM1 and HMM2), the bi-gram model (MM3) and the superposition model (SP) of a part of speech bi-gram and morpheme bi-gram[5], using Japanese RWC corpus. The results showed that the nested joint probability model (NEST) achieved the highest performance for open data condition: 95.8% precision rate against 94.5% of the superposition model and 94.0% of the established HMM2 model (Table 1). Also the model displayed only a small difference in performance between closed (training) data and open (variation) data condition, therefore we can claim that our model is more robust than the others.

Although the NEST model needs a large dataset of joint frequencies, we found that most of them are rejected by the statistical test and never used for the estimation. Therefore we also introduced a grid-pruning model of NEST that keeps only statistically significant interaction frequency data. This pruning model (DIET1) can save 94% of memory space for transition frequency data while maintaining almost the same accuracy as the NEST model (Table 1).

Table 1: Morphological Analysis Result of Nested Joint Probability Model (NEST) and the Grid Pruning Model (DIET1,2) Compared with HMMs and the Superposition Model

MODEL	Closed (Training) Data, N=6,916		Open (Validation) Data, N=9,015		Frequency Data Size**
	Recall Rate	Precision Rate	Recall Rate	Precision Rate	
HMM1	90.8	92.4	87.4	89.5	84
HMM2	94.9	96.3	92.0	94.0	737
MM3	97.2	97.4	68.2	75.0	31,656
SP	98.9	99.3	94.0	94.5	56,492
NEST	96.9	97.7	94.6	95.8	75,740
DIET1*	97.0	97.7	94.2	95.4	4,427
DIET2*	97.2	97.9	94.1	95.4	6,775

\* DIET1 : Grid Pruning based on the critical  $p=0.01\%$  ; DIET2 :  $p=5\%$ .

\*\* # of Frequency cell in transition matrix excluding morphemes dictionary.

## References

- Fujimoto, K., Inui, N., and Kotani, Y. 1998. Nested Joint Probability Model for Morphological Analysis. In Proceedings of the IASTED International Conference Artificial Intelligence and Soft Computing, Forthcoming.