論文 / 著書情報
Article / Book Information

論題(和文)	インドネシア語のコードスイッチング音声の認識
Title(English)	Recognition of Indonesian Code-Switching Speech
著者(和文)	ヨナタンアンディファジャルヌグラハ, 篠田浩一, 古井貞熙, 岩野公司
Authors(English)	Yonatan Andy Fajar Nugraha, Koichi Shinoda, Sadaoki Furui, Koji Iwano
出典(和文)	日本音響学会 2012年 春季研究発表会 講演論文集, Vol. , No. , pp. 75-76
Citation(English)	2012 Spring Meeting ASJ, Vol. , No. , pp. 75-76
発行日 / Pub. date	2012, 3

Recognition of Indonesian Code-Switching Speech*

☆ Yonatan Andy Fajar Nugraha[1], Koichi Shinoda[1], Sadaoki Furui[1] and Koji Iwano[2]
([1]Tokyo Institute of Technology, [2]Tokyo City University)

1 Code-Switching

Code-switching is a phenomenon where alternation occurs between two or more languages within a single conversation. This can be done either inter-sententially or intra-sententially. Inter-sentential code-switching means that the switching occurs across sentence boundaries, and intra-sentential code-switching implies that the switching occurs within a single sentence, for instance in:

> "This dress is very *kawaii* (cute), and it is only *ni sen en* (two thousand yen)".

In the linguistic field of study, grammatical aspects of code-switching have been a great interest since there is no clear grammatical rules for code-switching [2]. This issue becomes a problem in speech recognition on building the language model of code-switching speech. The data sparseness of written code-switching leads to unpracticality in building a statistical language model of code-switching.

Previous works on the code-switching problem covered acoustical aspects, mainly by performing language identification [5] or phone modeling [8]. Although those approaches can yield reasonable improvements, the grammatical aspects are still questionable.In this paper we present our approach of language modeling on code-switching. We propose a combined language model as a way to embed contextual information from monolingual models into a code-switching model.

2 Code-Switching in Indonesian Language

Indonesian language has a large number of borrowed words from other languages. Thus, when code-switching occurs, many variations appear. For instance, when someone code-switches between Indonesian and English in a sentence, regardless of speaker's proficiency in English, the same English word which has a pair of its transliterated word in Indonesian, might be pronounced variedly.

As an example, Indonesian word *komputer* (pronounced */k o m p u t e r/*) is borrowed from English word "computer" (*/k ah m p y uw t er/*). If we take a sample of a code-switching sentence "*Saya belajar* (I study) computer science", the word "computer" here is connected with another English word "science" hence it should be spoken with English pronunciation. Nevertheless, there are cases in which Indonesian pronunciation is used.

3 Combined Language Model

To cover the problem mentioned in Section 2, a contextual knowledge obtained from languages constructing the code-switching speech can help estimating the output word sequence. Based on this idea, we propose a new approach by combining monolingual language models to construct the code-switching language model.

The language models are combined by WFST composition operation:

$$LM_{com} = LM_1 \circ LM_2.$$

LM_1 refers to a task dependent grammar-based language model constructed from a code-switching corpus and LM_2 refers to the language model built from independent monolingual text corpus that follows the following formulation:

$$P(W) = P(W_1) \prod_{i=2}^{N} \tilde{P}(W_i|W_{i-1}) \qquad (1)$$

where

$$\tilde{P}(W_i|W_{i-1}) = \begin{cases} P(W_i) & \text{if } L_i \neq L_{i-1}, \\ P(W_i|W_{i-1}) & \text{if } L_i = L_{i-1}. \end{cases}$$

$P(W)$ represents the probability of the word sequence W where each W_i is one of the N words in the sequence. L_i refers to the language of word W_i.

4 Experiments

4.1 Setup

The acoustic model used in these experiments is Indonesian-English bilingual acoustic model trained from the Indonesian LVCSR corpus [4] and the WSJ

*インドネシア語のコードスイッチング音声の認識
ヨナタン アンディ ファジャル ヌグラハ [1]、岩野公司 [2]、篠田浩一 [1]、古井貞熙 [1]([1] 東工大, [2] 東京都市大)

Table 1 Test Set

Test Data	Language	No.of utterances
ID	Mono Ind.	1064
EN	Mono Eng.	1216
BI	CS Ind-Eng.	2128

corpus [7]. 65 phone classes including 26 Indonesian phones, 34 English phones, and 5 clustered phones sharing the same pronunciations were used to train the acoustic model.

The grammar-based language model LM_1 were constructed based on observations on approximately 200 queries of computer voice commands. Each rule combines the possibility of switching between Indonesian and English words. For LM_2, we used standard bigram language model trained from approximately 700 sentences from each language.

WFST cascades for the acoustic model and language models were constructed with Transducersaurus toolkit [6] and optimized with tools from OpenFST [1] library. Finally, T^3 decoder [3] was used for the decoding process.

The test sets contain monolingual Indonesian (ID), English (EN), and intra-sentential code-switching (BI) utterances of computer voice commands that were not included in the training. The data were collected from 10 male and 9 female Indonesian native speakers. Configuration for the test sets is described in Table 1.

4.2 Results

We compared the performance of each test set using grammar-based code-switching language model LM_1, monolingual language model LM_2, and the proposed combined language model LM_{com}. Table 2 shows the results of the experiments.

In each type of language models, experiments on ID data gave the best performance among other test data. This is due to the fact that all the speakers are native Indonesian thus they brought non-native English utterances in EN and BI test sets. Since the English corpus used to train the acoustic model is only from native speakers, these non-native English speeches cannot be well handled.

For every test data, we found that the experiments with LM_2 delivered better performances than LM_1, and our proposed method outperformed the others. We noticed that in LM_1, since we considered all switching possibilities, a lot of deviations occurred thus often produced meaningless sentences. We also observed, particularly in EN test data, there were many errors in word structures.

Table 2 Word Error Rate (%)

Test Data	LM_1	LM_2	LM_{com}
ID	8.6	6.7	5.6
EN	19.0	14.6	10.5
BI	12.3	11.5	10.0

On the other hand, LM_2 gave more significant improvement on monolingual data compared to the BI test data. However, it had a number of insertion errors caused by taking other contextual words from the bigram. We found that our proposed method can effectively prune unnecessary search paths generated from LM_2 thus yielded further improvement to the recognition results.

Across all test data, the proposed method gave 5.32% relative improvement to the LM_1 and 2.52% relative improvement to the LM_2.

5 Conclusion

We proposed a combined language model approach to handle code-switching utterances. We tested it on computer voice commands which include monolingual utterances of Indonesian and English, and intra-sentential code-switching utterances spoken by native Indonesian speakers. We showed that the proposed method is effective for all type of data.

For future work, we would investigate the effectiveness of our method on other languages. We would also explore possibilities of combining more than two languages.

References

[1] C. Allauzen *et al.*, CIAA'07, 11-23, 2007.

[2] K. F. Cantone, "Code-Switching in Bilingual Children", chapter 4, Springer, 2007.

[3] P. R. Dixon *et al.*, ASRU 2007, 443-448, 2007.

[4] D. P. Lestari *et al.*, Indonesian Scientific Conference in Japan, 17-22, 2006.

[5] T. Niesler and D. Willett, MULTILING-2006, paper 004, 2006.

[6] J. R. Novak *et al.*, Interspeech 2011, 1537-1540, 2011.

[7] D. B. Paul and J. M. Baker, HLT'91, 357-362, 1992.

[8] Q. Zhang *et al.*, Journal of Information Science and Engineering, vol. 26, 1491–1507, 2010.