CODE SWITCH LANGUAGE MODELING WITH FUNCTIONAL HEAD CONSTRAINT

Ying Li, Pascale Fung

Department of Electronic and Computer Engineering
The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong
eewing@ee.ust.hk, pascale@ece.ust.hk

ABSTRACT

In this paper, we propose for the first time to incorporate the linguistically well-known Functional Head Constraint into a code switch language model for speech recognition. Under this constraint, code switch cannot occur between the functional head and its complements. The constrained code switching language model is obtained by first expanding the search network with a translation model; and then using lattice-based parsing to restrict paths to those permissible under the Functional Head Constraint. We tested our system on two tasks of code switch speech recognition, namely lecture speech recognition and lunch conversation recognition. Our system reduces word error rates (WER) from a baseline mixed language model by 3.72% relative in the first task; and by 5.85% in the second task. It reduces WER from an interpolated language model by 2.51% in the first task; and by 4.57% in the second task. All results are statistically significant. In addition, our method reduces WER for both the matrix language and the embedded language.

Index Terms— Mixed language speech recognition, code switch language modeling

1. INTRODUCTION

Code switching is a common linguistic phenomenon among multilingual speakers. It is particularly prevalent among the large population of speakers of Spanish/English, Hindi/English, Chinese/English, Arabic/French, etc. Multilingual people often mix words or phrases in the principal language and code switch to the other language when they speak. The principal language of the speech is called the matrix language (ML), and the foreign language is the embedded language (EL) [1].

Lack of transcribed code switch training data is a major obstacle faced by speech recognition systems. Consequently, language modeling of code switch speech has been a major bottle neck to system development.

A baseline approach to code switch language modeling is to use an interpolated model between the monolingual language models of the matrix and embedded languages. This approach poses very little constraint on where languages can be switched between each other. Another approach is to detect the code switch boundaries of the mixed language speech by means of phonetic or acoustic modeling, then recognize the speech segment by using the corresponding monolingual language model [2, 3, 4, 5]. The major and often fatal weakness of this approach is that language identification at the switch points is often erroneous and leads to irrecoverable recognition errors down the pipeline.

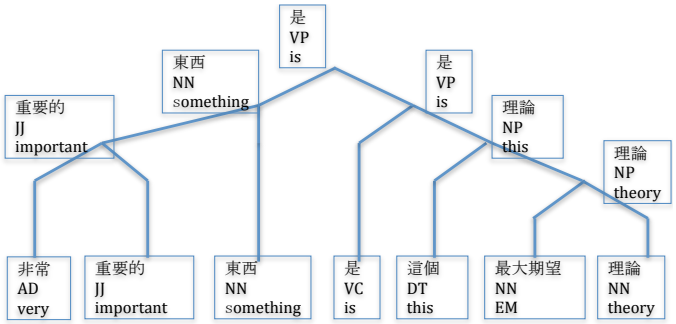
To provide more generality, linguists have long discovered and debated various constraints on code switching speech. Among them, the Functional Head Constraint [6, 7, 8], along with the Equivalence Constraint [12], are the best-known grammatical constraints. For code switching, they both provide grammatical rules that constrain where in a sentence code switch cannot happen (instead of postulating where and why code switch can happen). Functional Head Constraint is known to be more restrictive than the Equivalence Constraint. Under this constraint, code switch cannot occur between a functional head and its complements. In this paper, we propose to incorporate the Functional Head Constraint to a code switch language model for speech recognition.

Based on this, we use a two-pass decoding procedure to recognize the code switch speech. In the first pass, the code switch speech is decoded using the composition of the acoustic model, language model in the matrix language and a code switch language model. The code switching language model is trained using parallel text in the matrix and embedded languages. In the second pass, a lattice-based parser is used to obtain the syntactic structure to subject to the Functional Head Constraint. The code switch language model is then combined with a monolingual language model in the matrix language for final recognition.

2. RELATION TO PRIOR WORK

A common approach of code switch language model is to interpolate between the matrix and embedded languages, trained separately from monolingual texts [9, 10, 11]. However, this approach does not assume any syntactic constraint that has been found in the code switching speech.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE



Hypotheses: 非常重要的是這個 EM 理論.
非常重要的是這個 EM theory.
非常重要的東西是 this EM theory.
非常重要的東西 is this EM theory.
非常重要的 something is this EM theory. (not permissible)
.
.
.

Fig. 1. A Functional Head Constraint example.

We previously implemented an approximation of the Equivalence Constraint [12, 13, 14] in a code switch language model by using word order inversion constraint from machine translation [15, 16, 17].

The work presented here focuses on implementing the full Functional Head Constraint, which is found to be more restrictive than the Equivalence Constraint, by using a lattice parser with a weighted finite state transducer. The lattice-based parsing approach presented in the prior works [18] is used to parse the word lattice generated from the first decoding step.

3. FUNCTIONAL HEAD CONSTRAINT

Functional Head Constraint is a well-known observation made by linguists [6, 7, 8] regarding code switch. The Functional Head Constraint stipulates that "the language feature of the complement f-selected by a functional head, like all other relevant features, must match the corresponding feature of that functional head" [8]. This means that "code-switching cannot occur between a functional head (a complementizer, a determiner, an inflection, etc.) and its complement (sentence, noun-phrase, verb-phrase)" [1].

Our work for the first time propose a method to incorporate this linguistically well-know constraint into code switch language modeling.

We illustrate the Functional Head Constraint in Figure 1. Functional heads are the roots of the subtrees: complements are the subtrees. actual words are at the leave nodes. According to the Functional Head Constraint, all the leaves of a given subtree must be in either the matrix language or the embedded language.

[1] http://en.wikipedia.org/wiki/Code-switching

For instance, the third word "東西/something" is the head of the constituent "非常/very 重要的/important 東西/something". According to the Functional Head Constraint, it is not permissible to code switch between the three constituent words. More precisely, the language of the constituent is constrained to be the same as the language of the headword.

4. CODE SWITCH LANGUAGE MODELING WITH FUNCTIONAL HEAD CONSTRAINT

Each input speech is decoded to generate a word lattice using the acoustic and language models. A tree structure is obtained by parsing the word lattice using the lattice-based syntactic parser. The lattice is expanded by allowing the subtree of the parse tree to be code-switched to the embedded language according to the Functional Head Constraint.

4.1. Code Switch Language Modeling in a WFST Framework

We propose to integrate a code switch acoustic model and the code switch language model with Functional Head Constraint in a WFST framework.

Suppose X denotes the observed code switch speech vector, a hypothesis transcript is as follows:

$$\begin{aligned}
\hat{v}_1^I &= \arg\max_{v_1^I} P(v_1^J|X) \\
&= \arg\max_{v_1^I} P(X|v_1^I)P(v_1^I) \\
&= \arg\max_{v_1^I} P(X|v_1^I)\sum_{w_1^J} P(v_1^I|w_1^J)P(W_1^J) \\
&\cong \arg\max_{v_1^I} P(X|v_1^I)P(v_1^I|w_1^J)P(w_1^J) \qquad (1)
\end{aligned}$$

where $P(X|v_1^I)$ is the acoustic model and $P(v_1^I)$ is the language model in the mixed language.

As code switch text data is scarce, our code switch language model is obtained from a translation model $P(v_1^I|w_1^J)$ from the matrix language to the mixed language, and the language model in the matrix language $P(w_1^J)$.

Instead of word-to-word translation independently, the transduction of the context dependent lexicon transfer is constrained by previous words:

$$
\begin{aligned}
P(v_1^I|w_1^J) &= \prod_{i=1}^{I} P(v_i|v_1^{i-1}, w_1^i) \\
&\cong \prod_{i=1}^{I} P(v_{i-n+1}^{i-1}|w_{i-n+1}^{i}) \\
&= \prod_{i=1}^{I} \frac{P(v_i, w_i|v_{i-n+1}^{i-1}, w_{i-n+1}^{i-1})}{P(w_i|v_{i-n+1}^{i-1}, w_{i-n+1}^{i-1})} \\
&= \prod_{i=1}^{I} \frac{P(v_i, w_i|v_{i-n+1}^{i-1}, w_{i-n+1}^{i-1})}{P(w_i|\sum_{v_i} v_{i-n+1}^{i-1}, w_{i-n+1}^{i-1})} \quad (2)
\end{aligned}
$$

We propose to use a weighted finite state transducer framework incorporating the bilingual acoustic model P, the context model C, the lexicon L, and the grammar G into a C-level search network

$$N = P \circ C \circ L \circ G \quad (3)$$

The H-level search network is composed of the state model H the universal phoneme model P, the context model C, the lexicon L, and the grammar G

$$N = H \circ P \circ C \circ L \circ G \quad (4)$$

The C-level requires less memory then the H-level search network. For mixed language speech recognition, the grammar is code switching language model G_{CS}. The output of the recognition result is in the mixed language after projection $\pi(G_{CS})$:

$$N = P \circ C \circ L \circ \pi(G_{CS}) \quad (5)$$

In our previous system, the code switching language model with inversion constraint was modeled as

$$
\begin{aligned}
P(v_1^I|w_1^J) = & \sum_{\tilde{v}_1^L, c_1^L, r_1^K, \tilde{w}_1^K} P(\tilde{w}_1^K|w_1^J) \cdot P(r_1^K|\tilde{w}_1^K, w_1^J) \\
& \cdot P(c_1^L, r_1^K, \tilde{w}_1^K, w_1^J) \\
& \cdot P(\tilde{v}_1^K|c_1^L, r_1^K, \tilde{w}_1^K, w_1^J) \\
& \cdot P(v_1^I|\tilde{v}_1^K, r_1^K, \tilde{w}_1^K, w_1^J) \quad (6)
\end{aligned}
$$

where $P(\tilde{w}_1^K|w_1^J)$ segments the words into phrases, $P(r_1^K|\tilde{w}_1^K, w_1^J)$ estimates the probabilities of phrase permutations, $P(\tilde{v}_1^K|c_1^L, r_1^K, \tilde{w}_1^K, w_1^J)$ translates a chunk of words into the embedded language and $P(\tilde{v}_1^K|c_1^L, r_1^K, \tilde{w}_1^K, w_1^J)$ is reconstruct the chunks to words. A word sequence in the matrix language w_1^J is segmented into phrases, $\tilde{w}_1^K$; and $\tilde{v}_1^K$ is a chunk sequence in mixed language.

In this paper, the WFST implementation to obtain the code switch language model G_{CS} is proposed as follows:

$$G_{cs} = T \circ G \quad (7)$$

where T is the translation model

$$P(\tilde{v}_1^L|w_1^J) = \prod_{l=1}^{L} P_l(\tilde{v}_l|w_l) \quad (8)$$

$P_l(\tilde{v}_l|w_l)$ is the probability of w_l translated into $\tilde{v}_l$. The words are translated under the Functional Head Constraint.

4.2. Functional Head Constraint by Lattice-based Parsing

Lattice-based parsing is used to expand the word lattice according to the Functional Head Constraint. A Probabilistic Context-Free Grammar (PCFG) parser is trained on Penn Treebank data. The PCFG parser is generalized to take the lattice generated by the recognizer as the input. All the nodes of the word-lattice are ordered by increasing depth. The CYK table is obtained by associating the arcs with their start and end states in the lattice instead of their sentence position and initialized all the cells in the table corresponding to the arcs. The remaining procedure of CYK is unchanged.

After the parse tree is obtained, we recursively enumerate all its subtree. Each subtree is able to code-switch to the embedded language with a translation probability $P_l(\tilde{v}_l|w_l)$.

4.3. Decoding by Translation

In our work, a two-pass decoding is proposed to recognize code switch speech. The first decoding pass composes of the transducer of the universal phoneme model P, the transducer C from context-dependent phones to context-independent phones, the lexicon transducer L which maps context-independent phone sequences to word strings and the transducer of the language model G.

$$ASR_1 = P \circ C \circ L \circ G \quad (9)$$

The language model G_{CS} of the transducer in the second pass is improved from G by composing with the translation model $P_l(\tilde{v}_l|w_l)$. Finally, the recognition transducer is optimized by determination and minimization operations.

$$ASR_2 = P \circ C \circ min(det(L \circ min(det(\pi(G_{CS}))))) \quad (10)$$

5. EXPERIMENTS

5.1. Experimental Setup

The bilingual acoustic model is trained from 160 hours of GALE Phase 1 Chinese broadcast conversation, 40 hours of GALE Phase 1 English broadcast conversation, and 3 hours of nonnative English data recorded in house. 39-dimensional MFCC features analyzed at a 10msec frame rate with a 25msec window size are used. The pronunciation dictionary is obtained by modifying Mandarin and English dictionaries using the phone set consisting of 21 Mandarin standard

initials, 37 Mandarin finals, 6 zero initials and 6 extended English phones. The acoustic models are state-clustered crossword tri-phone Hidden Markov Models with 16 Gaussian mixture and reconstructed by decision tree tying [10].

We also use two Chinese-English code switching speech databases - namely, 20 hours of lecture speech corpus (Data 1) and 3 hours of lunch conversation corpus (Data 2). 1037 utterances of Data 1 are used as the test set (Test 1). 280 utterances of Data 2 are used as the test set (Test 2).

Transcriptions of 18 hours of Data 1 are trained as a baseline mixed language model for Test 1. Chinese speech conference papers, power point slides and web data are used to train a baseline Chinese language model for the lecture speech domain (LM 1). Transcriptions of 2 hours of Data 2 are used as the baseline mixed language model in Test 2. The GALE Phase 1 Chinese conversational speech transcriptions are used to train a Chinese language model (LM 2). GALE Phase 1 English conversational speech transcriptions are used to train the English language model (LM 3).

To train the bilingual translation model, the Chinese Gale Phrase 1 conversational speech transcriptions are used to generate a bilingual corpus using machine translation.

An interpolated language model for the lecture speech and lunch conversation domain is trained from interpolating LM 1 with LM 3 and LM 2 with LM 3 respectively.

Adapted language models are trained by adapting the interpolated models with code switch data under the maximum likelihood criterion using the EM algorithm [19].

5.2. Experimental Results

Table 1 and 2 shows the word error rates (WERs) of experiments on the code switching lecture speech and lunch conversation test sets. Our proposed code switching language model with Functional Head Constraints reduces the WERs in the baseline mixed language model by 3.72% relative on Test 1; and 5.85% on Test 2. Our method also reduces WER by 2.51% relative compared to the adapted language model on Test 1; and by 4.57% on Test 2. Moreover, our proposed method also reduces WER by 5.47% compared to the interpolated model on Test 2. All the WER reductions are statistically significant at 99%. 35 out of 2080 and 12 out of 570 code switching instances are corrected by the proposed method over Test 1 and Test 2.

For our reference, we also compare the performance of using Functional Head Constraint to that of using inversion constraint in our previous work and found that the present model reduces WER by 0.85% on Test 2 but gives no improvement on Test 1. We hypothesize that since Test 1 has mostly Chinese words, the proposed method is not as advantageous compared to our previous work. Another future direction is for us to improve the lattice parser as we believe it will lead to further improvement on the final result of our proposed method.

Table 1. *Our proposed system outperforms the baselines in terms of WER on the lecture speech*

	Mandarin	English	Mixed
MixedLM	34.41%	39.16%	35.17%(-3.72%)
InterpolatedLM	34.11%	40.28%	35.10%(-3.73%)
AdaptedLM	35.11%	38.41%	34.73%(-2.51%)
CodeSwitchingLM +Inversion	32.76%	37.00%	33.44%(+1.26%)
CodeSwitchingLM +FunctionalHead	**33.27%**	**36.94%**	**33.86%**

Table 2. *Our proposed system outperforms the baselines in terms of WER on the lunch conversation*

	Mandarin	English	Mixed
MixedLM	46.4%	48.55%	46.83%(-5.85%)
InterpolatedLM	46.04%	49.04%	46.64%(-5.46%)
AdaptedLM	46.64%	48.39%	46.20%(-4.57%)
CodeSwitchingLM +Inversion	43.85%	46.97%	44.47%(-0.85%)
CodeSwitchingLM +FunctionalHead	**43.24%**	**46.27%**	**43.89%**

6. CONCLUSION

In this paper we describe a first ever method of incorporating the linguistically well-known Functional Head Constraint in code switch speech into a code switch language model. Functional Head Constraint means that the complements of a subtree must switch language with its head word. We propose a weighted finite state transduction based framework to incorporate the acoustic model, the matrix language model, and the translation model for final decoding.

The translation model expands the matrix language model into a bilingual language model. Lattice-based parsing is used to provide the syntactic structure of the matrix language. Matrix words at the leave nodes of the syntax tree are permitted to switch to the embedded language if the switch does not violate the Functional Head Constraint. This reduces the permissible search paths from those expanded by the bilingual language model.

We tested our system on two tasks of code switched speech recognition, namely on lecture speech recognition and on lunch conversation recognition. Our system reduces WERs from a baseline mixed language model by 3.72% relative in the first task; and by 5.85% in the second task: it reduces WERs from an adapted language model by 2.51% in the first task; and by 4.57% in the second task. All results are statistically significant. In addition, our method reduces word error rates for both the matrix language and the embedded language.

7. REFERENCES

[1] F. Coulmas, *The handbook of sociolinguistics*, vol. 4, Wiley-Blackwell, 1998.

[2] N.T. Vu, D.C. Lyu, J. Weiner, D. Telaar, T. Schlippe, F. Blaicher, E.S. Chng, T. Schultz, and H. Li, "A first speech recognition system for mandarin-english code-switch conversational speech," 2012.

[3] J.Y.C. Chan, PC Ching, T. Lee, and H.M. Meng, "Detection of Language Boundary in Code-switching utterances by Bi-phone Probabilities," in *Chinese Spoken Language Processing, 2004 International Symposium on*. IEEE, 2005, pp. 293–296.

[4] C.J. Shia, Y.H. Chiu, J.H. Hsieh, and C.H. Wu, "Language boundary detection and identification of mixed-language speech based on map estimation," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*. IEEE, 2004, vol. 1, pp. I–381.

[5] D.C. Lyu and R.Y. Lyu, "Language identification on code-switching utterances using multiple cues," in *Ninth Annual Conference of the International Speech Communication Association*, 2008.

[6] Shahrzad Mahootian, *A null theory of codeswitching*, Ph.D. thesis, Northwestern University, 1993.

[7] Hedi M Belazi, Edward J Rubin, and Almeida Jacqueline Toribio, "Code switching and x-bar theory: The functional head constraint," *Linguistic inquiry*, pp. 221–237, 1994.

[8] Steven Paul Abney, *The English noun phrase in its sentential aspect*, Ph.D. thesis, Massachusetts Institute of Technology, 1987.

[9] K. Bhuvanagiri and S. Kopparapu, "An approach to mixed language automatic speech recognition," in *Oriental COCOSDA, Kathmandu, Nepal*, 2010.

[10] Y. Li, P. Fung, P. Xu, and Y. Liu, "Asymmetric acoustic modeling of mixed language speech," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 5004–5007.

[11] D. Imseng, H. Bourlard, M. Magimai-Doss, and J. Dines, "Language dependent universal phoneme posterior estimation for mixed language speech recognition," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 5012–5015.

[12] E. Woolford, "Bilingual code-switching and syntactic theory," *Linguistic Inquiry*, vol. 14, no. 3, pp. 520–536, 1983.

[13] J. MacSwan, "13 code-switching and grammatical theory," *The Handbook of Bilingualism and Multilingualism*, p. 323, 2012.

[14] S. Poplack and D. Sankoff, "A formal grammar for code-switching," *Papers in Linguistics: International Journal of Human Communication*, vol. 14, pp. 3–45, 1980.

[15] Ying Li and Pascale Fung, "Code-switch language model with inversion constraints for mixed language speech recognition.," in *COLING*, 2012, pp. 1671–1680.

[16] Ying Li and Pascale Fung, "Improved mixed language speech recognition using asymmetric acoustic model and language model with code-switch inversion constraints," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 7368–7372.

[17] Ying Li and Pascale Fung, "Language modeling for mixed language speech recognition using weighted phrase extraction," *Interspeech*, 2013.

[18] Jean-Cédric Chappelier, Martin Rajman, Ramón Aragüés, Antoine Rozenknop, et al., "Lattice parsing for speech recognition," in *Proc. of 6ème conférence sur le Traitement Automatique du Langage Naturel (TALN 99)*, 1999, pp. 95–104.

[19] Renato DeMori and Marcello Federico, "Language model adaptation," in *Computational models of speech pattern processing*, pp. 280–303. Springer, 1999.