基于条件随机场的泰语音节切分方法

赵世瑜 线岩团 郭剑毅 余正涛 洪玄贵 王红斌
(昆明理工大学信息工程与自动化学院 昆明 650500)

摘 要 音节是泰语构词和读音的基本单位,泰语音节切分对泰语词法分析、语音合成、语音识别研究具有重要意义。结合泰语音节构成特点,提出基于条件随机场(Conditional Random Fields)的泰语音节切分方法。该方法结合泰语字母类别和字母位置定义特征,采用条件随机场对泰语句子中的字母进行序列标注,实现泰语音节切分。在InterBEST 2009泰语语料的基础上,标注了泰语音节切分语料。针对该语料的实验表明,该方法能有效利用字母类别和字母位置信息实现泰语音节切分,其准确率、召回率和F值分别达到了99.115%、99.284%和99.199%。

关键词 泰语字母特征,泰语音节,音节切分,条件随机场

中图法分类号 TP391.1 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.3.010

Thai Syllable Segmentation Based on Conditional Random Fields

ZHAO Shi-yu XIAN Yan-tuan GUO Jian-yi YU Zheng-tao HONG Xuan-gui WANG Hong-bin
(Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China)

Abstract Syllable is the basic unit of word-formation and pronunciation of Thai. Thai syllable segmentation is significant to lexical analysis, speech synthesis and speech recognition. Combined with the characteristics of Thai syllables, Thai syllable segmentation method based CRFs (Conditional Random Fields) was proposed. In order to achieve Thai syllable segmentation, the algorithm not only combines the Thai alphabet categories and letter position to define features, but also employs CRFs for letters in Thai sentence to do sequence labeling. In this paper, Thai syllable segmentation corpus was marked on the basis of InterBEST 2009. Experiments for the corpus demonstrate the method can effectively achieve Thai syllable segmentation by adopting the category and location information of alphabetical letters, and the va-lues of precision, recall and F reach 99.115%, 99.284% and 99.199%.

Keywords Thai character feature, Thai syllable, Syllable segmentation, Conditional random fields

1 引言

音节切分是语音合成、语音识别的基础。在音节特性显著的语言中,语音识别通常选择音节作为识别的基本单元[1],泰语语音识别中也广泛使用了音节信息[2]。在语音合成(TTS)中,音节特征显著的语言也选择音节作为合成的基本单元[3]。音节切分可以作为词法分析的特征来辅助词法分析,提高词法分析的准确率。

泰语属于音位文字类型,主要由元音字母、辅音字母和声调符号组成,是音节特征显著的语言。泰语词汇由音节构成,包括单音节词、双音节词和多音节词。泰国本土泰语的基本词汇中很大一部分是单音节词。音节一般由元音(V)字母、辅音(C)字母和声调(T)3部分组成;但由于第一声调没有声调标识符,所以有的音节只由一个辅音(C)字母和一个元音(V)字母构成,一般构成形式为C^TV、CC^TV、CVC^T、$CCVC^T$等,如泰语句子**เราได้รับความช่วยเหลือจากอาจารย์**(我们得到老师的帮助。)泰语中除缩写词中使用“.”和部分对话句子中可能用“?、!”标点外,句子之间都使用空格作为分割符。若使用“-”作为句子中音节切分的分割标志,其对应的音节切分结果为“**เรา-ได้-รับ-ความ-ช่วย-เหลือ-จาก-อา-จารย์**”。句子**นี่คือเก้าอี้ค่ะ**(这是椅子。)由音节“**นี่-คือ-เก้า-อี้-ค่ะ**”组成。词**หนังสือ**(书)属于双音节词,由两个音节“**หนัง-สือ**”组成。

目前针对泰语分词的研究相对较多,也比较成熟,而针对泰语音节切分的研究甚少,基本没有针对泰语音节切分方面的研究。20世纪80年代泰国就开始对泰语的音节进行分析[4-6]。2002年,Wirote Aroonmanakun在研究泰语分词时,使用200多条规则实现了泰语音节的切分并取得了很好的效果[7]。但基于规则的泰语音节切分存在规则复杂、只有语言

到稿日期:2015-03-20 返修日期:2015-06-17 本文受国家自然科学基金:面向互联网的泰语-汉语双语语料获取及对齐方法研究(61363044),国家自然科学基金:面向汉语-泰语跨语言新闻事件检索方法研究(61462054),云南省教育厅重点项目:汉语-泰语跨语言新闻事件检索中的相似度计算研究(2014Z021)资助。

赵世瑜(1989-),男,硕士生,主要研究方向为自然语言处理,E-mail:shiyuzhaocn@gmail.com;**线岩团**(1981-),男,讲师,主要研究方向为信息检索、自然语言处理,E-mail:yantuan.xian@gmail.com(通信作者);**郭剑毅**(1964-),女,教授,主要研究方向为自然语言处理、机器学习,E-mail:gjade86@hotmail.com;**余正涛**(1970-),男,教授,主要研究方向为信息检索、自然语言处理,E-mail:ztyu@hotmail.com;**洪玄贵**(1990-),男,硕士生,主要研究方向为自然语言处理,E-mail:Meawhunter@gmail.com;**王红斌**(1983-),男,讲师,主要研究方向为信息检索、分布/并行计算机系统,E-mail:whbin2007@126.com。

学专家才能掌握规则的书写、规则量大时规则之间可能出现冲突、音节切分速度慢等问题。本文利用泰语字母构成泰语音节的特点，结合泰语字母及字母在音节中的位置特征，使用条件随机场模型对泰语音节进行分析，实现将泰语句子切分成音节，并在 InterBEST 2009 泰语分词语料的基础上标注了泰语音节切分语料，对本文使用的方法进行了验证。

2 条件随机场模型

条件随机场(CRFs)模型用于序列标注问题时，其核心思想是利用无向图理论学习训练数据，使序列标注的结果在整个观察序列上达到全局最优。条件随机场模型可以使用非独立的、复杂的和重叠的特征进行训练，同时克服了传统的隐马尔科夫模型(HMM)和最大熵马尔科夫模型(MEMM)的标记偏置等问题[8]。

条件随机场是一种基于无向图的条件概率判别式模型，它从训练数据中学习，使得序列标注的条件概率在整个序列上的概率最大化。线性链式条件随机场模型是一种简单而易用的模型，对于给定参数 $\Lambda=(\lambda_1,\lambda_2,\lambda_3,\cdots,\lambda_n)$ 的线性链式条件随机场模型，在给定输入序列 $X=x_1,\cdots,x_T$ 上，其对应的状态序列 $Y=y_1,\cdots,y_T$ 的条件概率为：

$$P_\Lambda(Y|X)=\frac{1}{Z_\Lambda(X)}\exp(\sum_{t=1}^{T}\sum_{k=1}^{K}\lambda_k f_k(y_{t-1},y_t,x,t))$$

其中，$f_k(y_{t-1},y_t,X,t)$ 为二值特征函数，λ_k 为该特征函数的权重；$Z_\Lambda(x)$ 是使所有状态序列概率和为 1 的归一化因子，其形式如下：

$$Z_\Lambda(X)=\sum_{Y}\exp(\sum_{t=1}^{T}\sum_{k=1}^{K}\lambda_k f_k(y_{t-1},y_t,x,t))$$

输入序列 $X=x_1\cdots x_T$ 的概率最大化标记序列为：

$$Y^*=\arg\max P_\Lambda(Y|X)$$

解码时广泛使用 Viterbi 算法进行。

3 基于条件随机场的泰语音节切分

泰语音节切分的主要任务是将输入的泰语句子切分为音节组合的形式。本文将句子中音节的切分转化为序列标注问题，使用条件随机场模型对泰语字母进行序列标注，从而达到泰语音节切分的目的。

3.1 条件随机场音节切分算法

设 C 为需要进行音节切分的泰语句子($C\in\{c_1,c_2,c_3,\cdots,c_n\}$，其中 c_n 表示泰语句子 C 中的第 n 个字符)；S 和 L 分别为泰语句子 C 上的音节序列和组成音节的泰语识别结果标记序列，其中 $L=\{l_1,l_2,l_3,\cdots,l_n\}$，$l_n$ 表示泰语句子 C 中的第 n 个泰语字符在音节中的位置标记，且 $l_n\in\{B,I\}$，而 $S=\{s_1,s_2,s_3,\cdots,s_m\}$ 表示句子 C 中的音节序列，s_m 表示句子 C 中的第 m 个音节。对待进行音节切分的泰语句子 C 中的每一个字母 c_n 进行序列标记，将泰语句子切分为音节 S 的组合形式。那么泰语音节切分任务就是找到一种标注，使得联合概率 $P(L|C)$ 最大化。即：

$$\hat{L}=\arg\max_{L} P(L|C) \tag{1}$$

根据条件随机场模型原理，条件概率 $P(L|C)$ 为：

$$P(L|C)=\frac{1}{N_L(C)}\exp(\sum_{i=1}^{n}\sum_{k=1}^{K_L}\lambda_k f_k(l_{i-1},l_i,c,i)) \tag{2}$$

式中，n 表示句子 C 中泰语字母的数目，λ_k 为条件随机场模型的参数，f_k 为条件随机场模型的特征函数，$N_L(C)$ 为归一化因子，其形式如下：

$$N_L(C)=\sum_{L}\exp(\sum_{i=1}^{n}\sum_{k=1}^{K_L}\lambda_k f_k(l_{i-1},l_i,c,i)) \tag{3}$$

我们把泰语音节切分转化为了基于条件随机场的泰语字母序列标注。假设标注集使用(B,I)，那么对泰语句子“เขาเป็นหัวหน้าห้อง”(他是班长。)使用条件随机场模型进行序列标注后得到的最好标注结果为“เ/B ข/I า/I เ/B ป/ ็/I น/I ห/B ั/I ว/I ห/B น/I ้/I า/I ห/B ้/I อ/I ง/I”，最后得到泰语音节的切分结果为“เขา|เป็น|หัว|หน้า|ห้อง|”。

3.2 标注集和特征选择

将泰语音节切分转化为序列标注问题，首先要定义标注集合。近年来在基于序列标注的中文分词系统中，广泛使用基于字在词中的位置来定义标注集，如 4 位标注集[9]或 6 位标注集[10]。由于泰语音节一般由多个字母组成，音节一般都由两个及两个以上的泰语字母组成，基本没有单字母成音节的情况，这里使用两位标注集(B,I)，B 表示泰语音节的首字母，I 表示泰语音节中除首字母之外的其它所有字母。

在特征选择方面，泰语由字母组成，字母又分为元音字母、辅音字母、声调字母及一些标志字母等，有的辅音字母不能出现在音节的结束位置，而有的元音字母不能出现在音节的首位置。因此选择泰语字母、字母类别和位置作为特征，如 c 表示辅音字母，v 表示元音字母、t 表示声调字母等。本文实验中使用 10 种音节切分的字母类别和位置特征，具体如表 1 所列。

表 1 泰语切分音节字母类别及位置特征

序号	标记	类别	字符
1	c	可以出现在音节末的辅音字母	ก ข ฃ ค ฆ ง จ ช ซ ญ ฎ ฏ ฑ ฒ ณ ด ต ถ ท ธ น บ ป พ ฟ ภ ม ย ร ล ว ศ ษ ส ฬ อ
2	n	不能出现在音节末的辅音字母	ฅ ฉ ฌ ผ ฝ ห ฮ ฤ ฦ
3	v	不能出现在音节起始位置的元音字母	ะ ั ิ ี ึ ื ุ ู า ำ ๅ
4	w	可以出现在音节开始位置的元音字母	เ แ โ ใ ไ
5	t	声调字母	่ ้ ๊ ๋
6	s	标志字母	ฯ ๆ ็ ์ ๏ ๚ ๛
7	d	数字	0-9 ๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
8	e	英语字母	a-zA-Z
9	q	引用字符	‘-’,“-”
10	o	其它字符	฿ ? !etc.

如果用 c_i 表示第 i 个泰语字母，$T(c_i)$ 表示条件随机场模型中第 i 个泰语字母 c_i 对应的类别，$L(c_i)$ 表示音节的长度信息，则特征模板如表 2 所列。

表 2 特征模板

条件随机场模型泰语音节切分特征模板
c_i，$i\in[-2,2]$且 $i\in Z$
c_ic_{i+1}，$i\in[-2,1]$且 $i\in Z$
$c_ic_{i+1}c_{i+2}$，$i\in[-2,0]$且 $i\in Z$
$T(c_i)$，$i\in[-2,2]$且 $i\in Z$
$T(c_i)T(c_{i+1})$，$i\in[-2,1]$且 $i\in Z$
$T(c_i)T(c_{i+1})T(c_{i+2})$，$i\in[-2,0]$且 $i\in Z$
$L(c_i)$，$i\in[-2,2]$且 $i\in Z$

表中 c_i 表示字母或音节特征，$T(c_i)$表示音节或字母的类别特征，$L(c_i)$表示音节或词中的字母或音节的长度信息，c_ic_{i+1}为字母或音节组合特征，$T(c_i)T(c_{i+1})$为字母或音节类别组合特征。

4 实验与分析

4.1 实验语料与评价标准

本文实验使用 2009 年 InterBEST 2009 泰语分词评测语料[11,12]。该语料分为 article、encyclopedia、news、novel 4 类，但该语料只能用作分词，并不能用于泰语音节切分。使用文献[7]中的规则将该分词语料切分为音节语料，然后对其结果进行人工校对，得到所需要的泰语音节切分语料。实验时分别从 4 类语料中随机抽取一部分作为测试语料，另一部分作为训练语料。实验语料分配如表 3 所列。

本文研究基于条件随机场模型的泰语音节切分，实验不使用其它资源，仅利用分词语料自身的资源对泰语音节切分性能进行评估，并以最终的准确率 P、召回率 R 及它们之间的调和平均值 $F(F=2PR/(P+R))$作为评价标准。

表 3 实验语料信息

类别	训练语料		测试语料	
	数据大小/MB	音节数	数据大小/MB	音节数
article	12.3	1443676	0.91	100537
encyclopedia	12.5	1390232	0.867	94291
news	18.7	2107488	1.16	134014
novel	16.0	1791859	1.42	154460
total	59.5	6733255	4.357	483302

为了验证实验数据中可能出现的偏移现象对结果的影响及对训练语料和测试语料选择的合理性，本文还将 4 类语料分为 10 份，其中 9 份作为训练语料，另 1 份作为测试语料进行 10 次交叉实验。

4.2 音节切分实验

在训练和测试实验中，根据本文的标注集和特征选择方法对 4 种类别的语料进行泰语音节切分实验。为了验证不同数量泰语字母特征对泰语音节切分性能的影响，实验时以泰语字母作为基本单位，构建不同泰语音节的 n-gram 进行实验（未特别说明时本文中所写的特征模板对应 5-gram）。表 4 为 4 种类别语料在全部训练集中学习到的分音节的准确率 P、召回率 R 和 F 值结果。

表 4 泰语音节切分实验结果

类别	3-gram			5-gram			7-gram		
	P(%)	R(%)	F(%)	P(%)	R(%)	F(%)	P(%)	R(%)	F(%)
article	97.687	97.893	97.790	99.121	99.354	99.237	99.171	99.371	99.271
encyclopedia	97.769	97.964	97.866	98.804	99.110	98.957	98.904	99.230	99.067
news	96.618	97.364	96.990	98.448	98.687	98.568	98.714	98.864	98.789
novel	97.989	98.238	98.113	99.095	99.481	99.288	99.149	99.455	99.301

表 4 中的实验结果显示，5-gram 时泰语音节切分的准确率已经达到了 99%，这主要是因为泰语中一个音节所包含的字母一般不超过 5 个，因此当 n-gram 达到 5 后，再增大 n-gram 对泰语音节切分准确率的改善已经不再明显。实验中在 4 个类别的语料上最后得到的准确率都达到或非常接近 99%，准确率最低的新闻类别在 5-gram 时也达到了 98.568%。实验中泰语音节切分获得了很高的准确率，其一是因为结合泰语音节构成特点，使用了泰语字母和泰语字母在音节中的位置作为特征；其二在于音节的形式相对固定，未登录音节出现的概率很小。音节切分错误一般出现在人名或一些直译的外来泰语词汇中的音节[7]，且在部分使用缩写形式泰语文章中，缩写没有完整的音节形式，如พล.ต.ต.ม.ร.ว.，这样的缩写不符合音节构成的特点，也可能出现音节切分错误。因为新闻类别语料中包含大量的人名及一些直译词汇，实验中新闻类别音节切分结果的准确率相对较低，也说明了泰语人名或一些直译外来泰语词汇中音节切分更容易出现错误，因此在泰语音节切分时新闻类别语料的准确率相对于其它类别偏低。

表 5 交叉实验结果

序号	训练音节数	测试音节数	P(%)	R(%)	F(%)
1	6476707	740400	99.230	99.262	99.246
2	6470596	746511	99.137	99.273	99.205
3	6535244	681863	99.027	99.248	99.137
4	6497852	719255	99.102	99.287	99.194
5	6478324	738783	99.170	99.333	99.251
6	6493736	723371	99.077	99.266	99.171
7	6508228	708879	99.240	99.339	99.290
8	6524309	692798	99.103	99.318	99.210
9	6492271	724836	99.126	99.331	99.228
10	6476696	740411	98.936	99.187	99.062
平均值			99.115	99.284	99.199

将 4 类语料分别分为 10 份，其中 9 份作为训练语料，剩下的 1 份作为测试语料，使用 5-gram 的特征模板进行 10 次交叉实验，以验证训练语料和测试语料选择的合理性并防止语料特征类别过于集中等情况下引起的泛化能力不强和过拟合等问题出现。表 5 显示了在 10 次实验中的泰语音节切分的实验效果。

交叉实验结果显示，10 次实验中，泰语音节切分的平均 F 值为 99.199%，而且每一次实验所得到的结果的 F 值都超过了 99%，说明不管以何种形式划分训练集和测试集，都获得了很好的泰语音节切分性能，表明本文使用的条件随机场模型泰语音节切分方法不仅获得了很好的泰语音节切分效果，而且性能稳定。

结束语 本文结合泰语字母类别和字母位置特征，使用条件随机场模型对泰语音节进行切分。实验结果证明该方法对泰语音节切分具有很好的效果，并且获得了稳定的泰语音节切分性能；本文的方法相对于基于规则的方法而言，不需要借助专业语言学知识编写复杂切分规则；另外，老挝语、缅甸语和柬埔寨语与泰语在构词上具有很多相似之处，因此本文研究也能为这些语言的相关研究提供借鉴。在进一步研究方面，考虑将泰语音节切分用于辅助泰语词法分析，寻求融入音节特征的泰语词法分析方法，研究音节切分对泰语词法分析结果、未登录词识别的影响。

参 考 文 献

[1] Yamamoto K, Nakagawa S. Comparison of syllab-le-based and phoneme-based DNN-HMM in Japane-se speech recognition[C]// 2014 International Conference Advanced Informatics: Concept, Theory and Application (ICAICTA). Bandung, 2014: 249-254

（下转第 83 页）

会影响聚类精度。这 3 个参数实际上主要是 r 有着决定作用。固定 c_1，c_2 的值，研究聚类精度随着 r 的变化会有怎样的变化。在 $c_1=2$，$c_4=1$，$c_3=0.001$ 情况下，实验结果如图 3 所示。从图 3 可以看出聚类精度并不是随 r 的增大而变大，这是由于 r 很大时，比如极端情形下，每个数据点的邻居都包括数据集中其余的数据点，在此情况下，数据点的逗留概率很大，则数据点的更新幅度很大，进而会产生聚类错误。然而在 r 很小时，数据点又几乎不进行移动。所以选择合适的 r 值很重要，我们将在下一步工作中对此进行深入研究。

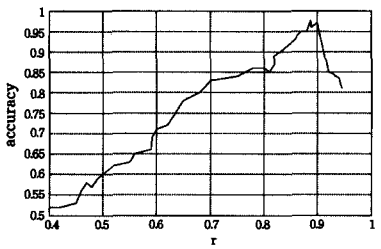

图 3　不同 r 的 Iris 的聚类精度

结束语　基于一维三态量子游走的“局部化”的特性，把数据点看作游走空间中的粒子，使得粒子执行若干步三态量子游走后，达到相似数据点聚集到一起的目的。首先，根据邻居集合和近邻集合构造粒子的游走空间；然后，根据定义的逗留概率构造了三态量子游走的硬币矩阵；最后，依据更新规则来实现数据点的移动。实验仿真结果证明了该算法的有效性。一维离散量子游走是最简单的量子游走模型，还有很多复杂且有趣的量子游走模式，如平面量子游走、散射量子游走等。在今后的研究中，我们将对这些量子游走进行研究，分析它们的特性，并尝试将其应用到聚类分析中。

参 考 文 献

[1] Shor P W. Algorithms for quantum computation : Discrete logarithms and factoring[C]// Proceedings of 35th Annual Symposium on Foundations of Computer Science, 1994. IEEE, 1994: 124-134

[2] Grover L K. Quantum mechanics helps in searching for a needle in a haystack[J]. Physical Review Letters, 1997, 79(2): 325-328

[3] Farhi E, Goldstone J, Gutmann S, et al. A quantum adiabatic evolution algorithm applied to random instances of an NP-complete problem [J]. Science, 2001, 292(5516): 472-475

[4] Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases[J]. AI magazine, 1996, 17 (3): 37-54

[5] SchuldM, Sinayskiy I, et al. An introduction to quantum machine learning[J]. Contemporary Physics, 2015, 56(2): 172-185

[6] Wang S H, Long G L. Big data and quantum computation[J]. Chinese Science Bulletin, 2015, 60(5/6): 499-508(in Chinese)
王书浩，龙桂鲁. 大数据与量子计算[J]. 科学通报(中文版)，2015, 60(5/6): 499-508

[7] Venegas-Andraca S E. Quantum walks: a comprehensive review [J]. Quantum Information Processing, 2012, 11(5): 1015-1106

[8] Childs A M. Universal computation by quantum walk[J]. Physical Review Letters, 2009, 102(18): 180501

[9] Li Q, He Y, Jiang J. A hybrid classical-quantum clustering algorithm based on quantum walks[J]. Quantum Information Processing, 2011, 10(1): 13-26

[10] Inui N, Konno N, Segawa E. One-dimensional three-state quantum walk[J]. Physical Review E, 2005, 72(5): 056112

[11] Erkan G. Language model-based document clustering using random walks[C]// Proceedings of the main conference on human language technology conference of the north American chapter of the association of computational linguistics. Association for Computational Linguistics, 2006: 479-486

[12] MacQueen J. Some methods for classification and analysis of multivariate observations[J]. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1967, 1 (14): 281-297

(上接第 56 页)

[2] Tangwongsan S, Phoophuangpairoj R. Boosting Thai Syllable Speech Recognition Using Acoustic Models Combination[C]// International Conference on Computer and Electrical Engineering(ICCEE 2008). 2008: 568-572

[3] Gu Hung-yan, Lai Ming-uen, Tsai Sung-Feng. Combining HMM Spectrum Models and ANN Pros-ody Models for Speech Synthesis of Syllable Prom-inent Languages[C]// 2010 7th International Symposium Chinese Spoken Language Processing (ISCSLP). Tainan, 2010: 451-454

[4] Thairatananond Y. Towards the Design of a Thai Text Syllable Analyzer [D]. Asian Institute of Technology, 1981

[5] Charnyapornpong S. A Thai syllable separation alg-orithm [D]. Asian Institute of Technology, 1983

[6] Poowarawan Y. Dictionary - based Thai syllable separathion [C]// Proceedings of the Ninth Electronics Engineering Conference. 1986

[7] Aroonmanakun W. Collocation and Thai Word Segmentation [C]// Proceedings of SNLP-Oriental Cocosda, 2002. 2002: 68-75

[8] Fferty J, McCallum A, Pereira F. Conditional random fields Probabilistie models for segmenting and labeling sequence data [C]// ICML2001. San Francisco: Morgan Kaufmann, 2001: 282-289

[9] Sproat R, Emerson T. The first international Chines-e word segmentation bakeoff[C] // 2nd SIGHAN Workshop on Chinese Language Processing. Morristown, NJ: ACL, 2003: 133-143

[10] Zhao Hai, Huang Chang-ning, Li Mu. An improved Chinese word segmentation system with conditional random field[C]// 5th SIGHAN Workshop on Chinese Language Processing. Morristown, NJ: ACL, 2006: 108-117

[11] Segmentation Guidelines for InterBEST 2009 Thai Word Segmentation: An international episode [EB/OL]. http://thailang.nectec. or. th/downloadcenter/index. php? option = com _ docman&-task=cat_view&-gid=43&-Itemid=61

[12] Boriboon M, et al. BEST Corpus Development and Analysis [C]// International Conference on Asian Language Processing, 2009 (IALP'09). 2009: 322-327