Algorithme de décodage de treillis selon le critère du coût moyen pour la reconnaissance de la parole

Antoine Rozenknop (1) et Marius Silaghi (2)
EPFL (DI-LIA) CH-1015 Lausanne (Suisse)
(1) Antoine.Rozenknop@epfl.ch, (2) Marius.Silaghi@epfl.ch

Résumé - Abstract

Les modèles de langage stochastiques utilisés pour la reconnaissance de la parole continue, ainsi que dans certains systèmes de traitement automatique de la langue, favorisent pour la plupart l'interprétation d'un signal par les phrases les plus courtes possibles, celles-ci étant par construction bien souvent affectées des coûts les plus bas. Cet article expose un algorithme permettant de répondre à ce problème en remplaçant le coût habituel affecté par le modèle de langage par sa moyenne sur la longueur de la phrase considérée. Cet algorithme est très général et peut être adapté aisément à de nombreux modèles de langage, y compris sur des tâches d'analyse syntaxique.

Stochastic language models used for continous speech recognition, and also in some Automated Language Processing systems, often favor the shortest interpretation of a signal, which are affected with the lowest costs by construction. To cope with this problem, this article presents an algorithm that allows the computation of the sequence with the lowest mean cost, in a very systematical way. This algorithm can easily be adapted to several kinds of language models, and to other tasks, such as syntactic analysis.

Mots-clefs/Keywords : Continuous speech recognition, Stochastic language models, Mean score.

1 Introduction

La reconnaissance de la parole continue à partir d'un signal acoustique est un problème d'une grande complexité du fait de la taille de l'espace de recherche des solutions. Afin de la rendre envisageable, il faut impérativement restreindre cet espace, par exemple en utilisant des modèles de langage. Cependant, même ainsi, le nombre de solutions correspondant à une réalisation acoustique proche du signal observé reste toujours très grand; une probabilisation *a priori* de l'espace de recherche est donc nécessaire (Murveit, Moore, 1990). Des exemples de tels modèles de langage probabilistes sont accessibles et bien étudiés, parfois dans un cadre de Traitement Automatique du Langage, mais présentent certains défauts s'ils sont utilisés sans adaptation préalable au problème de la reconnaissance de la parole. En particulier, les plus utilisés d'entre eux, qui reposent sur une modélisation paramétrique de processus stochastiques

(N-grams, grammaires stochastiques) affectent par construction des probabilités moindres aux phrases les plus longues, ce qui peut induire un fort biais lorsque le nombre de mots prononcés n'est pas connu à l'avance.

Une présentation succinte de l'utilisation de modèles de langages valués en reconnaissance de la parole est donnée dans la section 2. Nous y exposons aussi une idée tout-à-fait classique : elle consiste à utiliser en reconnaissance non pas le coût d'une phrase telle qu'un modèle de langage stochastique le définit (c'est-à-dire comme l'opposé du logarithme de sa probabilité), mais la moyenne de ce coût sur le nombre de mots de la phrase. Nous présentons ensuite dans la section 3 un algorithme original de décodage itératif, qui permet de déterminer les solutions de plus faible coût moyen, en s'appuyant sur les algorithmes existants de recherche de la solution de plus faible coût global. L'exposé de l'algorithme est suivi de sa preuve, ainsi que d'un petit exemple de déroulement, où l'on utilise une grammaire stochastique pour décoder un treillis de mots.

La mention de certains modèles de langage dans cet article a pour but de faire sentir l'intérêt qu'il y a à pouvoir extraire une phrase de coût *moyen* minimum. Il ne faut cependant en aucun cas y voir une limitation pour l'algorithme présenté, dont la grande force est justement son caractère très général.

2 Modèles de langage valués pour la reconnaissance de la parole

Nombre de systèmes de reconnaissance de la parole s'appuient sur des modèles de langage probabilistes pour sélectionner une séquence de mots parmi les différentes interprétations possibles d'un signal. Chaque interprétation M reçoit alors un "coût" acoustique $C_a(M)$ calculé par un module acoustique, ainsi qu'un "coût" linguistique $C_l(M)$ calculé à l'aide du modèle de langage probabiliste considéré; l'interprétation sélectionnée est celle ayant le coût total $C(M) = C_a(M) + C_l(M)$ minimal.

Or, les modèles de langage les plus utilisés sont des modèles génératifs stochastiques, qui *produisent* des séquences de mots par une succession d'étapes aléatoires. Les modèles de N-grams et les grammaires stochastiques hors-context (SCFG) en sont de bons exemples. L'intérêt de tels modèles est double : d'une part, leurs paramètres sont facilement obtensibles à partir de bases d'exemples; d'autre part, l'existence d'algorithmes efficaces autorisent leur exploitation effective pour le décodage de signaux de parole (Mteer, Jelinek, 1993). Mais ils présentent aussi un inconvénient majeur : le coût d'une séquence, calculé comme l'opposé du logarithme de sa probabilité d'être produite par le modèle, croît rapidement avec le nombre d'étapes du processus de production, donc avec le nombre de mots qui la composent. Ainsi ces modèles considèrent-ils que les hypothèses les plus courtes sont toujours les meilleures !

Pour pallier ce problème, on cherche en général à minimiser $C(M) - \beta|M|$ plutôt que $C(M)$ ($|M|$ est le nombre de mots de M, et β est une constante empiriquement déterminée).

Une autre idée naturelle est de chercher à minimiser le *coût moyen par mot* $\bar{C}(M) = C(M)/|M|$. C'est ce que permet de réaliser l'algorithme présenté dans la suite. Comme il utilise itérativement l'algorithme qui calcule $\mathrm{Argmin}_M\, C(M) - \beta|M|$, il ne requiert aucun espace mémoire supplémentaire; le nombre d'itérations nécessaires à l'obtention du résultat est le seul surcoût algo-

rithmique, et peut être majoré par $\max|M| - |M^0|$.

3 Algorithme de décodage itératif

3.1 Ingrédients

On dispose des éléments suivants :

- un ensemble E de phrases appartenant à un langage L, chaque phrase M étant constituée de $|M|$ mots,
- une fonction de coût C qui à chaque élément M de L associe un réel $C(M)$,
- un algorithme $\mathcal{A}(E, C, \beta)$ qui permet d'extraire :
  $\mathrm{Argmin}_{M\in E}\, C(M) - \beta|M|$ pour n'importe quel réel β.

3.2 Réalisation

L'algorithme $\mathcal{I}(\mathcal{A}, E, C)$ suivant permet de trouver une solution M^0 de $\mathrm{Argmin}_{M\in E}\, \bar{C}(M)$, où $\bar{C} = C(M)/|M|$:

1. Initialisation : on pose $\bar{C}(M_{-1}) = 0$.[1][2]
2. Itérations : on calcule $M_i = \mathrm{Argmin}_{M\in E}\, C(M) - \bar{C}(M_{i-1}) \cdot |M|$ en utilisant l'algorithme $\mathcal{A}(E, C, \bar{C}(M_{i-1}))$[3].
3. Critère d'arrêt : on cesse les itérations lorsque $|M_i| = |M_{i-1}|$. M_i est alors une solution du problème.

3.3 Preuve

Théorème 1 *L'algorithme $\mathcal{I}(\mathcal{A}, E, C)$ converge vers une solution $M^0 = \mathrm{Argmin}_{M\in E}\, \bar{C}(M)$ en un nombre d'itérations inférieur à $|M_1| - |M^0|$.*

Lemme 1 *Le coût moyen $\bar{C}(M_i)$ décroît strictement avec i, pour i supérieur à 0 et tant que $\bar{C}(M^0)$ n'a pas été atteint :*

$$\forall i \geq 0 \ \left[\bar{C}(M_i) > \bar{C}(M^0) \Rightarrow \bar{C}(M_i) > \bar{C}(M_{i+1})\right]$$

1. M_{-1} n'existant pas, ceci n'est qu'une convention d'écriture.
2. La valeur initiale de $\bar{C}(M_{-1})$ est sans importance pour la correction de l'algorithme. Le choix du 0 est arbitraire, et de fait, si l'on a une estimation *a priori* de la valeur de l'optimum $\bar{C}(M^0)$, le choix de cette estimation pour $\bar{C}(M_{-1})$ accélérera la convergence de l'algorithme par rapport au choix de la valeur 0.
3. i est l'indice de l'itération en cours, et vaut 0 pour la première itération.

Démonstration du lemme 1 :
Notons $C'_i(M) = C(M) - \bar{C}(M_i)|M|$.
Par définition de $\bar{C}$, on remarque immédiatement que : $C'_i(M) = |M|(\bar{C}(M) - \bar{C}(M_i))$.
L'algorithme $\mathcal{A}(E, C, \bar{C}(M_i))$ trouve une solution M_{i+1} qui minimise $C'_i(M)$, d'où :

$$\begin{aligned}
M_{i+1} = \mathcal{A}(E, C, \bar{C}(M_i)) \quad &\Rightarrow \quad M_{i+1} = \operatorname*{Argmin}_{M \in E} C'_i(M) \\
&\Rightarrow \quad C'_i(M_{i+1}) \leq C'_i(M^0) \\
&\Rightarrow \quad |M_{i+1}|(\bar{C}(M_{i+1}) - \bar{C}(M_i)) \leq |M^0|(\bar{C}(M^0) - \bar{C}(M_i)) \\
&\Rightarrow \quad \bar{C}(M_{i+1}) \leq \bar{C}(M_i) + \frac{|M^0|}{|M_{i+1}|}(\bar{C}(M^0) - \bar{C}(M_i)) \\
&\Rightarrow \quad \bar{C}(M_{i+1}) < \bar{C}(M_i)
\end{aligned}$$

La dernière ligne de la démonstration vient de l'hypothèse $\bar{C}(M_i) > \bar{C}(M^0)$, et de la stricte positivité de $|M|$ pour tout M appartement à E.

Lemme 2 *La taille $|M_i|$ des solutions successives décroît strictement pour i supérieur à 1 et tant que $\bar{C}(M^0)$ n'a pas été atteint :*

$$\forall i \geq 1 \ \left[\bar{C}(M_i) > \bar{C}(M^0) \Rightarrow |M_{i+1}| < |M_i|\right]$$

Démonstration du lemme 2 :
Pour $i \geq 1$,

$$\begin{aligned}
M_i = \mathcal{A}(E, C, \bar{C}(M_{i-1})) \quad &\Rightarrow \quad M_i = \operatorname*{Argmin}_{M \in E} C'_{i-1}(M) \\
&\Rightarrow \quad C'_{i-1}(M_{i+1}) \geq C'_{i-1}(M_i) \\
&\Rightarrow \quad |M_{i+1}|(\bar{C}(M_{i+1}) - \bar{C}(M_{i-1})) \geq |M_i|(\bar{C}(M_i) - \bar{C}(M_{i-1}))
\end{aligned}$$

Or d'après le lemme 1, $\bar{C}(M_i) > \bar{C}(M_{i+1})$, ce qui permet de minorer le second membre de l'inégalité précédente, et d'obtenir par transitivité :

$$|M_{i+1}|(\bar{C}(M_{i+1}) - \bar{C}(M_{i-1})) \geq |M_i|(\bar{C}(M_{i+1}) - \bar{C}(M_{i-1}))$$

Toujours d'après le lemme 1, $\bar{C}(M_{i-1}) > \bar{C}(M_i)$ (car $i - 1 \geq 0$), donc $\bar{C}(M_{i-1}) > \bar{C}(M_{i+1})$. On peut alors simplifier les deux membres de l'inégalité précédente, en la renversant, ce qui donne finalement :

$$|M_{i+1}| < |M_i|$$

Démonstration du théorème 1 :
Le lemme 2 montre que la taille des solutions M_i décroît strictement pour $i \geq 1$ tant que $\bar{C}(M_i) > \bar{C}(M^0)$. Comme cette taille est un entier strictement positif, elle cesse forcément de décroître pour un certain $i = i_f$, ce qui implique que $\bar{C}(M_{i_f}) = \bar{C}(M^0)$. L'algorithme atteint donc la solution du problème en un nombre fini d'itérations, et comme $|M_i|$ décroît strictement de $i = 1$ à $i = i_f - 1$, le nombre d'itérations i_f vérifie : $i_f \leq |M_1| - |M_{i_f}| = |M_1| - |M^0|$.

3.4 Exemple de déroulement

Afin d'illustrer l'algorithme itératif, nous présentons dans cette partie un petit exemple de décodage d'un treillis de mots à l'aide d'une grammaire stochastique. Les règles de la grammaire apparaissent dans la figure 1. Le treillis à décoder contient trois interprétations possibles : (1) `Célimène`, (2) `Céline mène` et (3) `C'est l'hymen`. À chacune correspond un arbre syntaxique, représenté dans la figure 2, avec son coût associé. On rappelle que le coût d'un arbre est la somme des coûts des règles qui le constituent, et que le coût moyen est cette même somme divisée par le nombre de feuilles de l'arbre.

Règle R		$P(R)$	$C(R)$	Règle R		$P(R)$	$C(R)$
R_{11}	S → NP	$1/3$	$0,477$	R_{23}	V → mène	$1/2$	$0,301$
R_{21}	S → NP V	$1/3$	$0,477$	R_{33}	V → est	$1/2$	$0,301$
R_{31}	S → P V A N	$1/3$	$0,477$	R_{35}	N → hymen	$1/15$	$1,176$
R_{12}	NP → Célimène	$1/2$	$0,301$	R_5	N → bateau	$14/15$	$0,030$
R_{22}	NP → Céline	$1/2$	$0,301$	R_4	A → un	$1/2$	$0,301$
R_{32}	P → c'	1	0	R_{34}	A → l'	$1/2$	$0,301$

FIG. 1 – *Règles de la grammaire avec leurs probabilités et leurs coûts. Le coût d'une règle vaut* $-\log P(R)$.

M	A_1	A_2	A_3
Interprétation	`Célimène`	`Céline mène`	`C'est l'hymen`
Règles	R_{11}, R_{12}	R_{21}, R_{22}, R_{23}	$R_{31}, R_{32}, R_{33}, R_{34}, R_{35}$
Probabilité	$1/6$	$1/12$	$1/180$
Coût	$0,778$	$1,079$	$2,255$
Coût moyen	$0,778$	$0,540$	$0,564$

FIG. 2 – *Coûts des différentes interprétations possibles.*

Les étapes de l'algorithme sont détaillées dans la figure 3, chaque ligne y représentant une itération, avec : (1) l'indice de l'itération, (2) le coût moyen de l'interprétation extraite lors de l'itération précédente, (3) le critère à minimiser, (4,5,6) les valeurs du critère pour les différentes interprétations possibles, (7) l'interprétation qui minimise le critère et (8) son coût moyen.

L'algorithme d'analyse syntaxique utilisé (Chappelier et al., 1999; Chappelier,Rajman,1998) peut extraire la solution qui minimise le critère $C'_i(M)$ en utilisant pour les règles terminales R_α ($\alpha \in \{12, 22, 23, 32, 33, 34, 35, 4, 5\}$) les coûts $C(R_\alpha) - \bar{C}(M_{i-1})$ à la place de $C(R_\alpha)$.

On peut remarquer que, comme prévu par les lemmes 1 et 2, le coût moyen des solutions successives diminue à partir de $i = 0$, et que leur taille diminue à partir de $i = 1$.

4 Conclusion

Dans cet article, nous avons décrit une classe de modèles de langage valués, utilisés pour la reconnaissance de la parole et permettant de décoder un signal en extrayant la phrase de coût minimal. Ces modèles ayant souvent la propriété de faire dépendre le coût d'une phrase de sa longueur, on en a dérivé un «meta-algorithme», aussi général que possible, qui extrait

Itération	$\bar{C}(M_{i-1})$	$C'_i(M)$	Valeurs de $C'_i(M)$			M_i	$\bar{C}(M_i)$
			A_1	A_2	A_3		
$i=0$	0	$C(M)$	$0,778$	$1,08$	$2,26$	A_1	$0,778$
$i=1$	$0,778$	$C(M)-0,778\|M\|$	0	$-0,477$	$-0,852$	A_3	$0,565$
$i=2$	$0,565$	$C(M)-0,565\|M\|$	$0,213$	$-0,051$	0	A_2	$0,540$
$i=3$	$0,540$	$C(M)-0,540\|M\|$	$0,238$	0	$0,1$	A_2	$0,540$

FIG. 3 – *Déroulement de l'algorithme itératif*

du signal une phrase de coût moyen minimal, et qui repose sur l'itération d'un algorithme spécifique au modèle de langage considéré. Sur une première expérience, on a constaté que le nombre d'itérations nécessaire avant la convergence reste très faible, même par rapport à sa valeur maximale théorique, ce qui rend ce «meta-algorithme» intéressant du point de vue de son efficacité. En revanche, la pertinence de l'utilisation du coût moyen comme critère d'extraction doit encore être évaluée pour d'autres modèles de langage stochastiques, et en fonction de l'application considérée.

Références

F.Itakura A.Ogawa, K.Takeda. "balancing acoustic and linguistic probabilities". *IEEE*, pages 181–184, 1998.

C. Chelba. *Exploiting Syntactic Structure for Natural Language Modeling*. PhD thesis, John Hopkins University, Baltimore, Maryland, 2000.

J.-M. Boite H. Bourlard, B. D'Hoore. Optimizing recognition and rejection performance in wordspotting systems. In *ICASSP'94*, volume I, pages 373–376, 1994.

H.Murveit and R.Moore. "integrating natural language constraints into hmm-based speech recognition". In *ICASSP'90*, pages 573–576, 1990.

C.-H. Lee E.R. Goodman J.G. Wilpon, L.R. Rabiner. Application of hidden markov models of keywords in unconstrained speech. In *ICASSP'89*, pages 254–257, 1989.

M. Eskénazi L.F. Lamel, J.-L. Gauvain. BREF, a large vocabulary spoken corpus for french. In *Eurospeech'91*, pages 505–508, 1991.

B.Juang L.Rabiner. *Fundamentals of Speech Recognition*. Prentice-Hall, 1993.

M.Meteer and J.R.Rohlicek. "statistical language modeling combining n-gram and context-free grammars". In *Proc.of ICASSP'93*, volume 2, pages 37–40, 1993.

M.C. Silaghi and H. Bourlard. "A new keyword spotting approach based on iterative dynamic programming." In *ICASSP'2000*, Istanbul, 2000.

J.-C.Chappelier, M.Rajman, R.Aragüés, A.Rozenknop. "Lattice Parsing for Speech Recognition" In *TALN'99*, pages 95–104, 1999.

J.-C.Chappelier, M.Rajman. "A generalized CYK algorithm for parsing stochastic CFG" In *TAPD'98*, pages 133–137, 1998.