Stochastic Language Design for Speech Recognition

Toru HITAKA,Yoichi TOMIURA and Natsuki ICHIMARU
Department of Computer Science and Communication Engineering
Kyushu University
6-10-1 Hakozaki, Higashi-ku, Fukuoka 810, JAPAN
E-mail : hitaka@lang.ai.kyushu-u.ac.jp
: tom@lang.ai.kyushu-u.ac.jp
: itimaru@lang.ai.kyushu-u.ac.jp

Stochastic Context Free Grammars are widely used as a convenient probabilistic language model for speech recognition. The stochastic grammar has parameters to be set accordingly to the set $\mathcal{F}_N$ of N sample sentences with grammatical structure. In previous studies of the parameter estimation, as known as most likelihood estimation, some evaluation function on parameters and on $\mathcal{F}$ is introduced first, then the values of parameters which maximize the value of the function are chosen. But this type of parameter estimation theory so far lacks something very important, that is consistency. Iff estimated values of parameters converge upon true values in probability when the number of samples becomes large, that estimation is called consistent. Our study obectives is to find a consistent parameter estimation for stochastic context free grammars. Our main results are as follows.

1. A stochastic grammar is called inconsistent, iff the sum of probabilities of all sentences is smaller than 1. In our study, it is proved that sample set of grammatical structures of sentences does not reflect on true values of parameters. This means that for an inconsistent grammar, the expected number of times a rewriting rule is applied is not proportionate to the true value of the parameter attached to the rule, so it is hard to come up with a consistent parameter estimation for inconsistent stochastic context free grammars. Therefore, characterization of consistent (or inconsistent) grammar is very important for our research. Existence of a inconsistent stochastic context free grammar was first introduced by T.L. Booth and J.D. Thompson in 1973 and they gave a sufficient condition on which a stochastic context free grammar is consistent. But this condition is proved in our research to be only a sufficient one, not a necessary one, and they offered no procedure to compute the sum of probabilities of all sentences. We solved almost all problems related to this subject, as shown below.

   (1) Characteristic Equations for which the sum of probabilities of all sentences is one of the solutions are introduced from the definition of stochastic context free grammars.

   (2) The sum of probabilities of all sentences is characterized as the smallest positive solution of the associated characteristic equations.

   (3) Asymptotic Expansion Formula for the smallest positive solution of characteristic equations is given, so that the sum of probabilities of all sentences can be computed with adequate accuracy.

   (4) All non-ill-formed stochastic regular grammars, in which every nonterminal symbol derives at least a string of terminal symbols and is derived in a sentencial form from the starting symbol S, are consistent.

2. Let δ, T and $c(\delta, T)$ be a rewriting rule, a grammatical structure (a tree) and the number of times δ is applied in T, respectively. We proved the expected value $\overline{c}(\delta)$, defined below, is proportionate to the true value $p(\delta)$ of the parameter δ only when the corresponding stochastic grammer is consistent. This finding suggests that it is difficult to come up with a consistent parameter estimation for inconsistent stochastic context free grammears.

$$\overline{c}(\delta) \stackrel{\text{def}}{=} \sum_{T \in \mathcal{T}} c(\delta, T) P_r(T)$$

where, $\mathcal{T}$ is the set of all grammatial structures and $P_r(T)$ is the occurrence probability of T.

Therefore, for all rewriting rules $\delta_{i1}, \delta_{i2}, \cdots, \delta_{im_i}$ whose left side nonterminal symbol is X_i,

$$p(\delta_{ij}) = \frac{\overline{c}(\delta_{ij})}{\overline{c}(\delta_{i1}) + \overline{c}(\delta_{i2}) + \cdots + \overline{c}(\delta_{im_i})} \quad (j = 1, 2, \cdots, m_i) \quad \cdots \quad (1)$$

Foumula (1) leads to the parameter estimate formula (2) defined in 4.

3. There are some consistent stochastic grammars for which $\overline{c}(\delta)$ becomes infinite. In this case, it is difficult to produce a consistent parameter estimate formula. Another obstacle to get a consistent parameter estimate is whether variance of $c(\delta, T)$ is finite or not, because if variance is infinite, law of large numbers can't be applied.

$$\text{variance of } c(\delta, T) \stackrel{\text{def}}{=} \sum_{T \in \mathcal{T}} (c(\delta, T) - \overline{c}(\delta))^2 P_r(T).$$

We produced fomulas to calculate $\overline{c}(\delta)$ and variance of $c(\delta, T)$ respectively and showed that variance of $c(\delta, T)$ is finite iff $\overline{c}(\delta)$ is finite. Then, we gave a necessary and sufficient condition for existance of $\overline{c}(\delta)$.

4. Finally, we come to a parameter esimate formula (2).

$$\widehat{p}(\delta_{ij}, \mathcal{F}_N) = \frac{\sum_{T \in \mathcal{F}_N} c(\delta_{ij}, T)}{\sum_{k=1}^{m_i} \sum_{T \in \mathcal{F}_N} c(\delta_{ik}, T)} \quad \cdots \quad (2)$$

Parameter estimate is called *consistent* iff for any small number $\varepsilon(>0)$, there exists a natural number N_0 and following inequality holds for any $N(>N_0)$

$$P_r\{|\widehat{p}(\delta_{ij}, \mathcal{F}_N) - p(\delta_{ij})| < \varepsilon\} > 1 - \varepsilon$$

Where $P_r\{|\widehat{p}(\delta_{ij}, \mathcal{F}_N) - p(\delta_{ij})| < \varepsilon\}$ is the occurrence probability of $\mathcal{F}_N$ for which $|\widehat{p}(\delta_{ij}, \mathcal{F}_N) - p(\delta_{ij})| < \varepsilon$.

Using law of large numbers, we proved that if the corresponding stochastic grammar is consistent and $\overline{c}(\delta)$ is finite, then formula (2) is a consistent parameter estimate.

Thanks to this program, we solved most of the problems we had aimed at. There is one important problem yet to solve. That is to come up with a consistent parameter estimate formula even for a consistent stochastic grammar whose $\overline{c}(\delta)$ is infinite. We already have a hint to solve this problem.

keywords : parameter estimation, consistent estimation, stochastic grammar