REVUE ROUMAINE DE LINGUISTIQUE

TOME XXXII • 1987

CAHIERS DE LINGUISTIQUE THÉORIQUE ET APPLIQUÉE

TOME XXIV • 1987 • N° 1
JANVIER—JUIN

TIRAGE À PART

EDITURA ACADEMIEI REPUBLICII SOCIALISTE ROMÂNIA

ON THE INFERENCE OF PROBABILISTIC REGULAR GRAMMARS

RĂZVAN ANDONIE

1. INTRODUCTION

There are many information-processing problems wherein a system must be designed to process information generated by a source with unknown characteristics. The first task that faced the system designer is to develop a model of the source. If the source generates a collection of finite-length output sequences, then it is possible to use a grammar to model the source. The search for a grammar that generates the sequences is called *inference* [8].

We shall consider the following problem : the inference of a probabilistic regular grammar G_P^0 for which $L(G_P^0) = M_0$, where M_0 is a finite probabilistic set of finite strings of symbols, and the equality concerns also the strings' probabilities. The obtaining of grammar G_P^0 is a major step towards general probabilistic regular grammatical inference (see [1, 2, 6, 8]), with important implications in learning theory and syntactic pattern recognition.

A first attempt to solve this problem was made by K.S.Fu [5]. He proposed a procedure to obtain first the finite-state probabilistic automaton and afterwards the corresponding probabilistic regular grammar. The transition probabilities of the automaton are computed by solving an equations system which also contains the normalization conditions.

Two algorithms for obtaining directly grammar G have been proposed by F. Maryanski and T. Booth [8], and respectively by R. Andonie [1, 2]. The present paper described in a simplified, condensed form the inference procedure from [1, 2] and compares it with the algorithm of F. Maryanski et al.

2. PRELIMINARY DEFINITIONS

In this section we settle the basic terminology and motivation, in particular those concerning probabilistic grammars.

Let $\sum$ be a finite set of terminals, N a finite set of nonterminals, S∈N the initial symbol, and R a finite set of rules. A *probabilistic grammar* is a context-free grammar $G_p = (\sum, N, R, S)$, with the production rules in R with the following form : $X_i \to B_{ij}$, $X_i \in N$, $B_{ij} \in (\sum \cup N)^*$ where the rule is applied with probability $p_{ij} (0 \leqslant p_{ij} \leqslant 1)$. The probability p_{ij} of applying any particular rule from R will be assumed to be independent of the sequence of previously applied rules. The grammar is *proper* if $\sum_{j=1}^{n_i} p_{ij} = 1$

for each nonterminal X_i, where n_i is the number of all the rules with X_i on the left side.

Let G be the *characteristic grammar* of G_p (obtained by deleting in G_p the probabilities from the rules). Assuming that G is unambiguous, we have:

$L(G_p) = \{(x, p(x)) \mid x \in \sum^*, S \overset{*}{\underset{G_p}{\Rightarrow}} x\}$, where $p(x)$ is the product of the probabilities of the rules in the derivation. The language $L(G_p)$ is *consistent* if $\sum_{x \in L(G)} p(x) = 1$, where $L(G)$ is the language generated by grammar G.

A ***probabilistic regular grammar*** is a probabilistic grammar for which the characteristic grammar is regular.

The ***formal derivative*** of a set $M \subseteq \sum^*$ is defined as the following set [3, 7]: $D_x M = \{y \mid xy \in M\}$, $x, y \in \sum^*$.

3. THE INFERENCE PROCESS

Let $\sum$ be a finite set of symbols, and M_0 a finite set of finite word over the alphabet $\sum$, $M_0 = \{(x, p(x)) \mid p(x) > 0\}$. We aim to transcribe in a more convenient formalism the inference algorithm given in [1, 2].

The grammar G_P^0 is defined as follows:

1. Take all of the not necessarily distinct formal derivatives of M_0, not equal to $\{\lambda\}$ or the null set Φ (λ stands for the empty word). Let $N = \{X_1, X_q\}$ be the set of these derivatives, and $X_1 = D_\lambda M_0 = S$.

2. R is the set of production rules with the following form:

a. $$X_i = D_{a_1 \ldots a_{r-1}} M_0 \rightarrow a_r D_{a_1 \ldots a_{r-1} a_r} M_0 = a_r X_j$$

$$\text{iff } D_{a_1 \ldots a_r} M_0 \neq \{\lambda\}, \neq \Phi$$

The associated rule probability is $p(X_j)/p(X_i)$, where $p(D_x M_0) = \sum_{\substack{xy \in M_0 \\ y \in \Sigma^+}} p(xy)$

b. $$X_i = D_{a_1 \ldots a_{r-1}} M_0 \rightarrow a_r$$

$$\text{iff } \lambda \in D_{a_1 \ldots a_{r-1} a_r} M_0$$

The associated rule probability is $p(a_1, \ldots a_r)/p(X_i)$

We obtain thus the probabilistic regular grammar $G_P^0(\sum, N, R, S)$ with the following properties (proved in [2]):

1. G_P^0 is proper.
2. The associated characteristic grammar G^0 is deterministic.
3. $L(G_P^0)$ is consistent.
4. The nonterminals are not merged.
5. $L(G_P^0) = M_0$.

The equality concerns also the probabilities of the strings iff $q = \sum_{(x, P(x)) \in M_0} p(x) = 1$.

The quantity q may be regarded as a *measure of confidence* for the inferred grammar. If $q < 1$, this means that there are some other words in M (which are not known) and their probability sum is $1 - q$. Grammar G_P^0 is inferred only from the known words in M_0.

H. Cramér [4] has demonstrated that a maximum likelihood estimation of the unknown parameters of a given distribution results in a minimum value for the chi-square statistic. It may be proved [6, 8] that the maximum likelihood estimate of the probability of a rule $X_i \to \beta_{ij}$ is :

$$\hat{p}_{ij} = \frac{n_{ij}}{N_i} \tag{1}$$

where N_i is the number of times nonterminal X_i appears in the generation of the strings of M_0 according to grammar G_P^0, and n_{ij} is the number of times rule $X_i \to \beta_{ij}$ is used in the above generation.

It is obvious that the probability of a rule in G_P^0 is equal to $\hat{P}_{ij}$. Hence, the given algorithm achieved a maximum likelihood estimation referring to the information contained in M_0.

This inference procedure has been implemented and tested in the framework of a syntactic pattern recognition algorithm [1, 2].

4. CONCLUDING REMARKS

P. Maryanski and T. Booth [8] proposed a similar inference procedure with the following differences :

1. The probabilities of the rules are estimated from relation (1).
2. The characteristic grammar is obtained with merged nonterminals due to the fact that only distinct formal derivatives are considered.

Essentially the same grammar is inferred, in spite of the fact that the two algorithms are different.

REFERENCES

1. ANDONIE, R. *Probabilistic Regular Grammatical Inference in Pattern Recognition.* (In Romanian), Proceedings of the Sixth Colloquium on Informatics, Iassy, 27–29 Oct. 1983, pp. 402–406.
2. ANDONIE, R. *Syntactic and Informational Aspects in Pattern Recognition.* Doctoral thesis (in Romanian), Univ. of Bucharest, Dept. of Mathematics and Informatics, Bucharest, 1984.
3. BOOTH, T. L., *Sequential Machines and Automata Theory.* Wiley, New York, 7167.
4. CRAMÉR, H. *Mathematical Models of Statistics.* Princeton Univ. Press, Princeton, NJ, 1946.
5. FU, K. S. *On Syntactic Pattern Recognition and Stochastic Languages.* In : *Frontiers of Pattern Recognition.* S. Watanabe (red.), Academic Press, New York 1972, pp. 113–138.
6. FU, K. S., BOOTH. T. L. *Grammatical Inference : Introduction and Survey.* IEEE Trans. on Syst., Man, and Cybernetics, vol SMC–5, 1975, 95–111, 409–423.
7. MARCUS, S. *Grammars and Finite Automata.* (In Romanian). Ed. Academiei, București, 1964.
8. MARYANSKI, F. J., BOOTH, T. L. *Inference of Finite-State Probabilistic Grammars.* IEEE Trans. on Computers, vol C–26, 1977, 521–526.

Research Institute for Computers Technique and Informatics
University of Brașov