Abstract. Formal translations have become of great interest for modeling some Pattern Recognition problems, but they require a stochastic extension in order to deal with noisy and distorted patterns. A Maximum Likelihood estimation has been recently developed for learning the statistical parameters of Stochastic Regular Syntax-Directed Translation Schemes. The goal of this paper is the study of estimation criteria in order to take into account the problem of sparse training data. In particular, these are the Maximum Mutual Information criterion and the Conditional Maximum Likelihood criterion. Some experimental results are reported to compare the three criteria.

1 Introduction

A translation is a process that maps strings from a given language (input language) to strings from another language (output language). Initially, translations were proposed in Syntactic Pattern Recognition as a framework for a fine presentation of error-correction models [10]. However, this formalism has recently become of great interest as a model in some practical Pattern Recognition problems in which the classification paradigm is not adequate [15], since the number of classes could be large or even infinite. In this case, the most general paradigm of Interpretation seems to be a better framework for problems of this type which can be tackled through formal translations. For example, many tasks in Automatic Speech Recognition can be viewed as simple translations from acoustic sequences to sublexical or lexical sequences (Acoustic-Phonetic Decoding) or from acoustic or lexical sequences to sequences of commands to a data-base management system or to a robot (Semantic Decoding). Other more complex applications of formal transducers are Language Translations (i.e. English to Spanish). Another interest in formal machines for translation comes from the fact that these machines can be learned automatically from examples [13]. In general, however, the application of formal machines to Syntactic Pattern Recognition

* Work partially supported by the Spanish CICYT under grant TIC95-0884-C04-01
needs a stochastic extension due to the noisy and distorted patterns under consideration which make the process of interpretation ambiguous [8]. Maximum Likelihood algorithms for automatically learning the statistical parameters of Stochastic Regular Syntax-Directed Translation Schemes from examples have recently been proposed [5]. Under this criterion, for each input string, the learning process only affects the correct output string. Under the Maximum Mutual Information criterion [3] [4], for each pair of strings, the learning process also affects all the output strings which are possible translations of an input string of the training pair as well as all the possible input strings whose translation is the output string of the training pair. Similarly, in the Conditional Maximum Likelihood [12] approach for each pair of strings, the learning process affects all possible output strings which are possible translations of the same input string of the training pair. These criteria can be particularly useful when the training data is sparse. Some experimental results are also reported in order to study the influence of the size of the training data over the percentage of the correctly translated sentences.

2 Stochastic Regular Syntax-Directed Translation Schema

A Stochastic Regular Syntax-Directed Translation Schema (SRT) is defined to be a system $T = (N, \Sigma, \Delta, R, S, P)$, where $N$ is a finite set of non-terminal symbols, $\Sigma$ and $\Delta$ are finite sets of input and output terminal symbols, respectively, and $S \in N$ is the starting symbol, $R$ is a set of rules of the form $A \rightarrow aB, zB$ or $A \rightarrow a, z$ for $A, B \in N$ $a \in \Sigma, z \in \Delta^*$ and $P : R \rightarrow [0, 1]$ is an assignment of probabilities to the rules such that the sum of the probabilities of all rules for rewriting $A$ is equal to one (proper SRT [10]).

A translation form for a translation $(x, y) \in \Sigma^* \times \Delta^*$ is a finite length sequence of rules $tf = (r_1, r_2, ..., r_n)$ such that: $tf : (S, S) \xrightarrow{(x, y)}$, where $(S, S) \xrightarrow{(x_1A_1, y_1B_1)} (x_1x_2A_2, y_1y_2B_2) \xrightarrow{...} (x, y)$ and the corresponding probability is

$$Pr(tf|\Phi(P)) = P(r_1)P(r_2)...P(r_n)$$

where $\Phi(P)$ defines the set of statistical parameters related to $P$.

For a translation form $tf$ from $(S, S)$ to $(x, y)$, output$(tf)$ and input$(tf)$ will denote $y$ and $x$, respectively.

A SRT is ambiguous, if for some $(x, y) \in \Sigma^* \times \Delta^*$, there is more than one translation form $tf$ such that $tf : (S, S) \xrightarrow{(x, y)}$. In this case, the probability of the translation $(x, y)$ is

$$Pr(x, y|\Phi(P)) = \sum_{\forall tf/input(tf)=x/output(tf)=y} Pr(tf|\Phi(P)) .$$

Alternatively, a Viterbi score for $(x, y)$ can be defined as