A PROBABILITY-THEORETICAL DECISION MODEL FOR THE AUTOMATIC CLASSIFICATION OF SIGNALS

W.J. BACKHAUSEN

In the area of acoustic phonetics, automatic speech recognition represents a central field of research. The problems to be solved there can be described as follows: the acoustic signals or more precisely the electric signals, which are generated by spoken speech units, for example phonemes or words, are to be automatically distributed into classes by means of an apparatus; each of the resulting classes has to represent accurately one of the meanings that are relevant to those speech communication processes to whose automatic processing this apparatus is to be applied. The assignment of a speech signal to one of these meaning-classes is thus to be interpreted as the identification of that meaning, which belongs to the corresponding speech unit.

In the following, a probability-theoretical decision model will be developed that will serve to solve this task. Let the functioning of a general model for an automatic identifier, that consists of an analyser A and a logical assigner Z, be described by means of the following transformations:

$$(1)\quad \tau(E) = \tau Z\,(\tau A\,(E)) = S$$

where

$$(2)\quad \tau A\,(E) = U$$

$$(3)\quad \tau Z\,(U) = S$$

U and S are finite sets such that $U_k \varepsilon U$, $k=1,\ldots,K$ stands for output values of measurement of the analyser A and $S_j \varepsilon S$, $j=1,\ldots,J+1$ represents the output signals of the assigner Z; for the infinite set E of input signals E_i, $i=1,2\ldots$, which are automatically to be classified, let us assume that a finite, complete segmentation $Z^{E}{}_1$ into disjunctive meaning-classes is defined:

$$(4)\quad Z^{E}{}_1 := \{a_1, \ldots a_j \ldots, a_J\}$$

Further let us assume that the criteria on which this division into classes is based is not translatable into methods of measurement.

Because of $cd(E) \gg cd(S)$, a further complete segmentation $Z^{E}{}_2$ of the set E into disjunctive classes β_j, $j=1,\ldots,J+1$ is defined by (1) as:

$$(5)\quad Z^{E}{}_2 := \{\beta_1, \ldots \beta_j \ldots, \beta_{J+1}\}$$

With the requirement that β_{J+1} as the class of undecidable cases should remain empty, an ideal identifier is characterized by the fact that the identification-classes β_j according to (5) are identical to the meaning-classes α_j resulting from (4); without the assistance of further information about the signals $E_i \varepsilon E$ to be classified, i.e., identified in their meaning, this is possible only if for the segmentation Z^{E_3} defined in (2) the following is true:

$$(6)\quad Z^{E_1} = Z^{E_3}$$

The task of the assigner would then be to produce a trivial biunique projection of Z^{E_3} onto S.

In all other cases additional information on the input signal set E would have to be taken into consideration for the construction of an automatic identifier. For the case, relatively free of preconditions, namely that the probability $p(\alpha_j)$ of occurrence of the various meanings α_j is known, i.e., that a probability measure (P_α) is defined over the segmentation Z^{E_1} — in matrix representation as below —

$$(7)\quad (P_\alpha)_{rs} = p(\alpha_r) \text{ for } r = s$$
$$= 0 \text{ for } r \neq s$$

then the following holds true: taking the results of a test series, which gives us the required probabilities $p(U_k/\alpha_j)$ for all α_j, that input signals of the meaning-class α_j are projected by the analyser A onto the set of measurement values U_k, then all of the information that is available for the construction of the assigner Z can be described by the following matrix equation (for an exact derivation cf. [1])

$$(8)\quad (P_\alpha/U) = (P_\alpha).(PU/_\alpha)'.\left[\sum_k \sum_j (E_{kj}).(P_\alpha).(PU/_\alpha)'.(E_{kk})\right]^{-1}$$

where

$$(9)\quad (P_\alpha/U)_{jk} = p(\alpha_j/U_k)$$
$$(10)\quad (PU/_\alpha)_{kj} = p(U_k/\alpha_j)$$
$$(11)\quad (E_{kj})_{lm} = 1 \text{ for } l=k \text{ and } m=j$$
$$= 0 \text{ otherwise}$$

This implies that the measure of the *a posteriori* probability (P_α/U), i.e., for the probability that the analyser output signal U_k was generated from an input signal E_i from the meaning-class α_j, can be determined by means of the measures for both of the *a priori* probabilities (P_α) and $(PU/_\alpha)$.

In order to take the relevance of correct and incorrect decisions in the automatic identification of signal meanings into account, a risk matrix $(C_{\beta,\alpha})$ is defined, which contains those cost factors that are to be considered in the assigning of a signal from the meaning-class α_j to the identification-class β_l. The multiplication of (8) and $(C_{\beta,\alpha})$

$$(12)\quad (C_{\beta,\alpha}).(P_\alpha/U) = (C_\beta/U)$$

results then in a matrix, which indicates the average costs that arise in dependence on a given analyser output signal U_k in the event that the corresponding input signal E_i belongs to one of the identification-classes β_j. The task of the assigner Z would then be to look for the minimum in the column of the matrix (C_β/U), which corresponds to a given output signal U_k of the analyser A, and to output the signal S_j, which is assigned to that row in which the minimum is found.

The automatic identifier for spoken numbers that was constructed on the basis of this model is reported on in /2/; the rate of error of this identifier is only slightly more than 1% when working with spoken digits.

Institut für Phonetik und Kommunikationsforschung
Bonn

REFERENCES

Backhausen, W.J.
1970 "Automatische Signalklassifikation durch ein wahrscheinlichkeitstheoretisches Entscheidungsmodell", *Phonetica* 22.
1971 "Untersuchungen zur automatischen Identifikation gesprochener Zahlwörter", *Phonetica* 24.